# LA-UR-22-21012

**Approved for public release; distribution is unlimited.**

| | |
|---|---|
| **Title:** | Toward Forecasting Geomagnetic Storms |
| **Author(s):** | Chase, Eve Adde<br>Larsen, Brian Arthur |
| **Intended for:** | Report |
| **Issued:** | 2022-02-09 (rev.1) |



# Toward Forecasting Geomagnetic Storms

**Eve A. Chase, Ph.D**
**Brian A. Larsen, Ph.D**

LA-UR-22-21012
February 8, 2022

Los Alamos
NATIONAL LABORATORY

# Contents



# Figures

## Contents



## Tables

## Acknowledgements

Template made by Tyler Borgwardt, edited by Tony H. Shin on 05/17/2021

# 1 Overview

We present a selection of attempts to infer geomagnetic storm properties from time series data. Initially, we unsuccessfully attempt to predict the duration of a geomagnetic storm based on data preceding the onset of a geomagnetic storm. Similar techniques are also used to attempt to predict the maximum depth of a geomagnetic storm. Additionally, we present a case study in Gaussian Process regression, where we are able to accurately interpolate between geomagnetic storm parameters, but are unable to predict future storm behavior. Lastly, we describe attempts at using random forest regression with sliding windows to infer geomagnetic storm behavior at future times. These results did not provide robust forecasts of geomagnetic storm behavior, but did occasionally correctly predict a storm within the margin of error. Techniques presented in this paper include machine learning classification and regression, Gaussian Process regression, and machine learning regression with sliding windows. Many of the attempts presented in this document did not result in meaningful predictions of geomagnetic storm behavior. We hope this document can be used as a guide for future readers attempting to predict geomagnetic storm properties.

## 2 Description of Dst Data

In this study, we leverage the OMNI data set [1] to study geomagnetic storms from January 1963 to July 2021. This data set includes multiple measurements as a function of time; although, we are specifically interested in the Dst value (Disturbance storm time index), which can indicate the presence of a geomagnetic storm [2]. Dst correlates to the strength of the ring current around Earth, caused by charged particles. Dst is an integer value proportional to Earth's magnetic field, such that Dst values are often reported in units of nT. Highly negative Dst values indicate the presence of a geomagnetic storm. For this study, we use the OMNI Dst data recorded at one-hour increments, displayed in Figure 2-1. We note that, in addition to Dst, several other time-varying values are presented in the OMNI dataset, including pressure and electric field measurements (see Section 5.2).

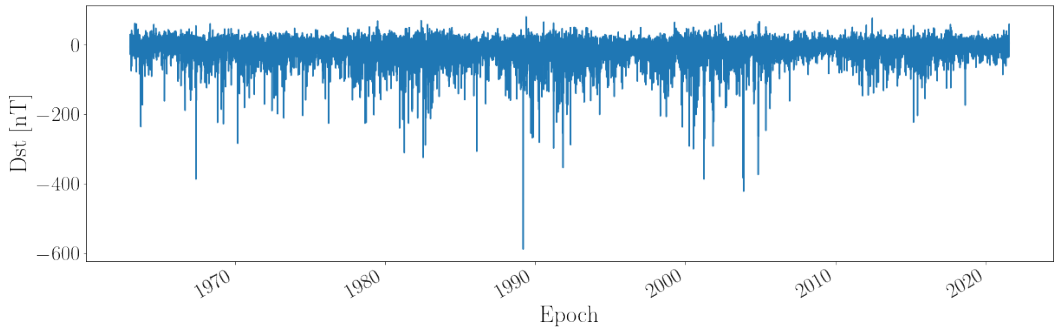



Figure 2-1: OMNI Dst data through July 2021.

We define the presence of a geomagnetic storm based on the Dst value, alone. We consider a storm start time of the first Dst value which falls below -45 nT. The storm is defined to end as soon as the Dst rises back above -8 nT. An example of a typical geomagnetic storm is displayed in Figure 2-2. We identify 1336 geomagnetic storms in the entire OMNI dataset.[1] These storms exhibit significant variation, including substantially different storm durations and varying minimum Dst values over the duration of the storm. In Figure 2-3, we present a selection of 20 storms from the OMNI dataset, superimposed with the first Dst value below -45 nT set to the zero-point. Significant variations in storm behavior are apparent.

In this work, we seek to predict geomagnetic storm duration and minimum Dst value given data preceding storm onset. Throughout this document, we present a variety of machine learning and statistically-motivated techniques in an effort to predict geomagnetic storm properties. Additional goals include predicting the overall storm morphology and predicting the likelihood of a storm occurring given observations several hours or days prior to the storm's onset. We specifically focus on predicting storm properties from Dst values alone, but eventually include auxiliary time-dependent measurements such as pressure and electric field in our prediction algorithms.

Prediction of storms remains a major challenge in the literature. The focus here upon minimum Dst and storm duration based on Dst alone is a limiting case meant to simplify the problem while providing space weather relevant information from easily available data. Many studies have worked

[1] We define storms where the Dst falls back below -45 nT within 10 hours of rising above -8 nT as a single multi-peaked storm, rather than two temporally nearby storms.

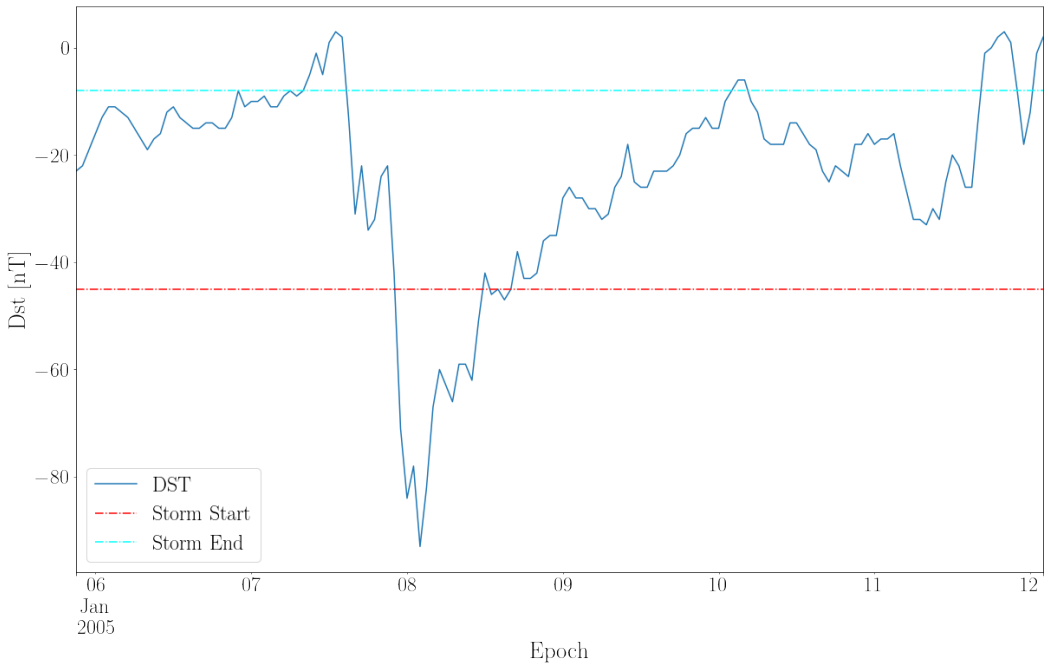



Figure 2-2: Example geomagnetic storm in OMNI data.

to predict Dst with decent success but none are quire relevant as they begin to bring in much more data and don't fit the mold we are attempting to address (i.e., see [6, 10, 11, 12, 13]).

Additionally, we refer the reader to the presentation in Reference [7] for an earlier example of this work, where the authors use $k$-dimensional ($k$-d) Tree Nearest Neighbors and Bayesian model-driven analysis to forecast storm behavior. Their $k$-d tree method involved loading all geomagnetic storm data into a $k$-d tree in 50 hour-long vectors. Then, these vectors were compared to a test set of Dst data. The three vectors most similar to the 50 hours of test data preceding a storm were then averaged to predict future storm behavior, as exemplified in Figure 2-4. These previous results demonstrated limited usefulness and serve as a motivation for the work presented in this document.

Results presented in this report are summarized in more detail on an internal LANL confluence page [3].

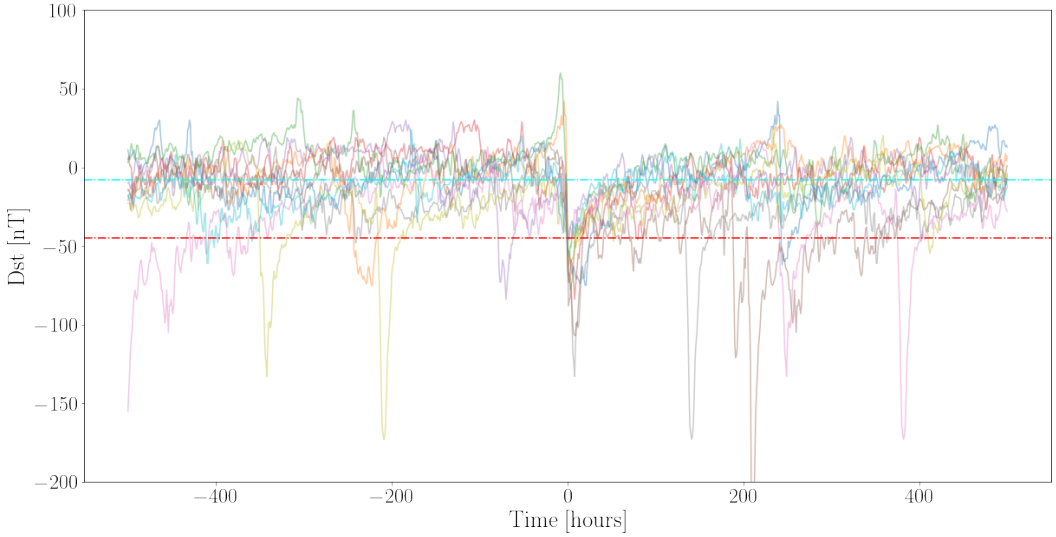



Figure 2-3: Twenty storms superimposed with time zero set to the storm onset. There is substantial variation in storm depth, length, and overall behavior.

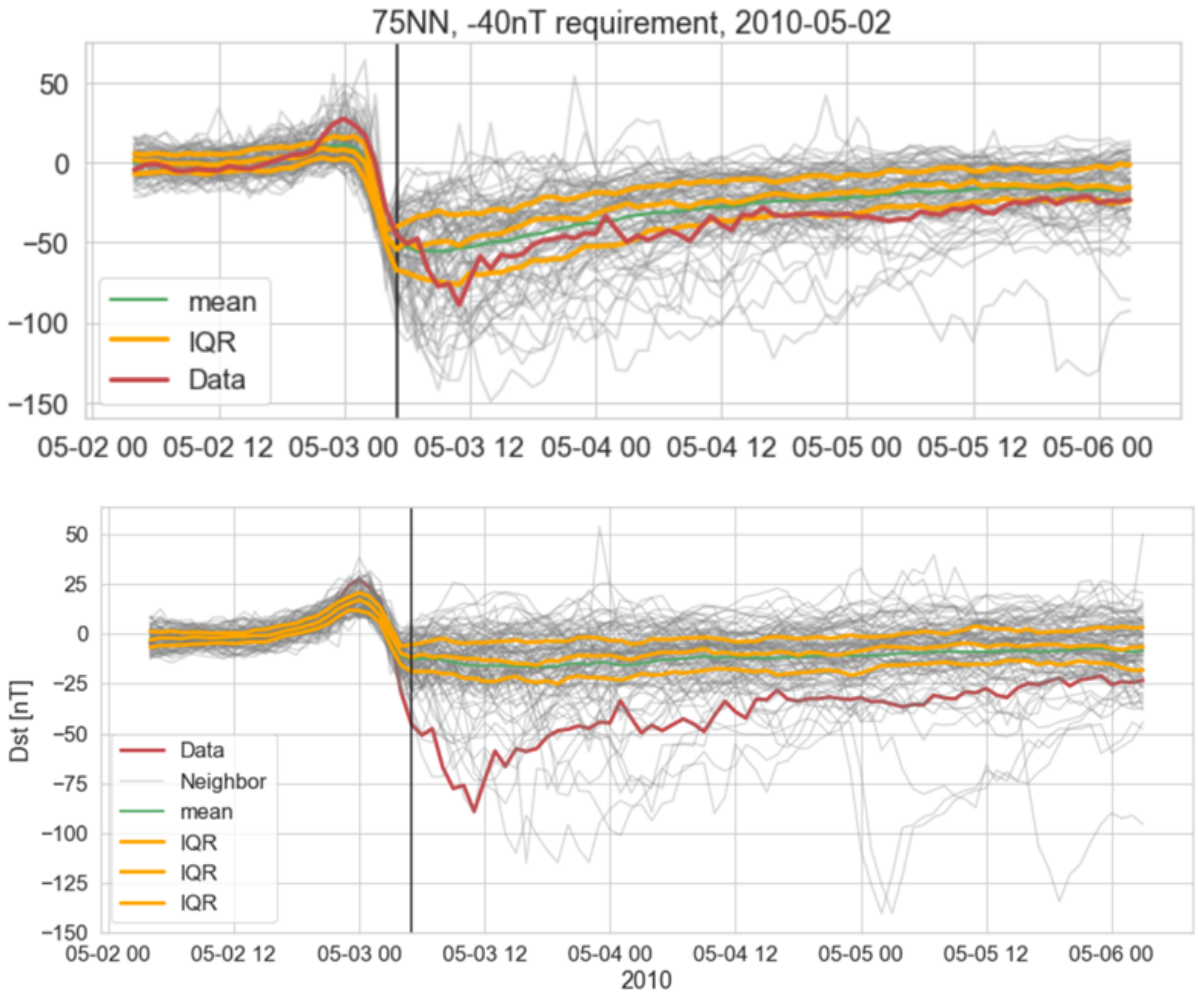



Figure 2-4: Two examples of using $k$-d tree to forecast Dst (from [7]).

# 3 Predicting Storm Duration

In this section, we present multiple attempts to infer storm length from Dst data, without any auxiliary time-dependent measurements. We employ multiple classification and regression techniques, available through the `scikit-learn` Python package [8].

## 3.1 Classification

In a classification problem, we label the data as belonging to one of a finite number of groups, based on some input data. In this example, we group geomagnetic storms into three classes, corresponding to the storm duration, as summarized in Table 3-1. Classes are designed so that the 1336 storms are roughly evenly distributed between the three classes.

Table 3-1: Storm duration classes.

| Storm Class | Class ID | Storm Duration |
|---|---|---|
| Short | 0 | Less than 54 hours |
| Medium | 1 | Between 54 and 111 hours |
| Long | 2 | More than 111 hours |

For all classification examples, we split the 1336 storms into testing and training sets, with 80% of samples used as training samples. We then train a random forest classifier [4] on the training set and use the testing set to evaluate the model's performance. Random forest classifiers are ensemble learning methods consisting of multiple decision tree classifiers, which work in unison to provide unbiased classification estimates. Decision trees classify instances by asking a series of binary (yes–no) questions of the input feature space (i.e. is the Dst at storm onset less than 80 nT?). Several hyperparameters dictate the decision tree's behavior including the maximum tree depth, minimum number of samples requires to further split a node, and maximum number of features considered in each branching decision. Hyperparameters of the classifier are selected using a grid search with three-fold cross validation.

Several metrics are presented to evaluate the classifier performance. All metrics are computed on the test set, after the random forest classifier is trained on the training set. We present confusion matrices, which correlates a storm's true class to the predicted class with the machine learning classifier. A highly accurate classifier would have strong support on the diagonal of a confusion matrix, with the majority of samples having their class accurately predicted. Additionally, we present an accuracy metric for each class, where accuracy is

$$\text{accuracy} = \frac{TP + TN}{N}, \tag{1}$$

where $N$ is the total number of samples, $TP$ is the true positive rate, and $TN$ is the true negative rate for a given class. Purity (also known as precision) corresponds to the fraction of samples predicted to belong to a class that are correctly identified:

$$\text{purity} = \frac{TP}{TP + FP}, \tag{2}$$

where $FP$ is the false positive rate. Likewise, completeness (also known as recall) refers to the fraction of samples with a true label for a given class that are correctly predicted to belong in that

class:

$$\text{completeness} = \frac{TP}{TP + FN}, \tag{3}$$

where $FN$ is the false negative rate. **A robust classifier has high values of accuracy, purity, and completeness, in tandem.**

All classifiers are optimized to predict the class corresponding to storm duration given some input data. We present multiple options for input data throughout the following subsections, including Dst values for various timesteps preceding the storm. Lastly, multiple input properties can be combined to produce additional classifiers.

### 3.1.1 Dst Value at Onset

In this example, we attempt to predict storm length based on the first Dst value of the storm (the first Dst value below -45nT). As demonstrated in Figure 3-1, there is little correlation between the Dst value and storm length. As a result, the classifier is unable to robustly predict the storm length. Overall, the training set is predicted with 42% accuracy while the testing set is predicted with 34% accuracy, suggesting mild overfitting. A confusion matrix for the fit is presented in Figure 3-2, indicating that regardless of true class, samples were predominately classified as "short." We present accuracies, purities, and completeness scores for each class in Table 3-2. This classifier has no utility for identifying storm length, regardless of the length of the storm, with an accuracy equivalent to random chance selection (∼33%).

Table 3-2: Classification metrics for Dst at onset.

| Storm Class | Accuracy | Purity | Completeness |
|---|---|---|---|
| Short | 0.44 | 0.34 | 0.65 |
| Medium | 0.61 | 0.25 | 0.10 |
| Long | 0.62 | 0.38 | 0.24 |

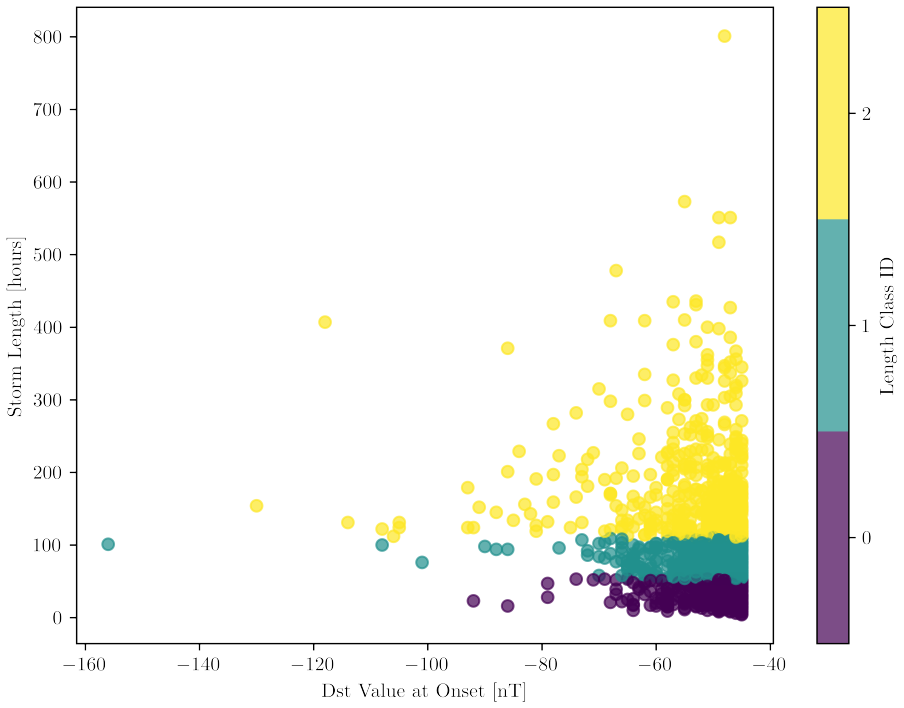



Figure 3-1: Scatterplot of storm length vs. Dst value at onset.

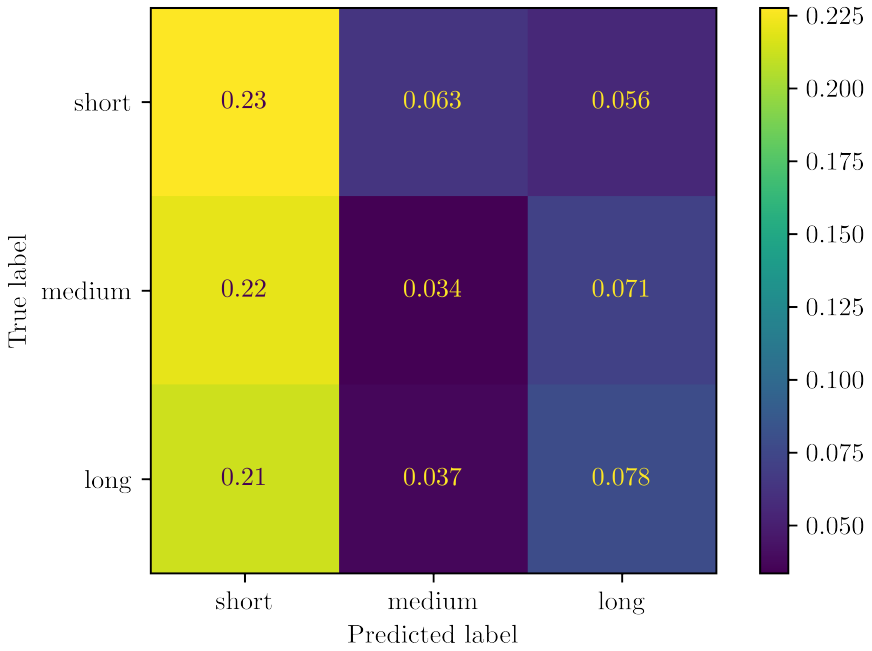



Figure 3-2: Confusion matrix for classification based on Dst value at onset. All confusion matrices in this document are normalized by the total number of testing samples.

### 3.1.2 Dst Slope at Onset

In this example, we attempt to predict storm length based on the slope at storm onset, computed from the difference between the Dst value at onset and Dst value one hour before onset. As demonstrated in Figure 3-3, there is little correlation between the slope and storm length. As a result, the classifier is unable to robustly predict the storm length. Overall, the training set is predicted with 43% accuracy while the testing set is predicted with 36% accuracy, suggesting mild overfitting. A confusion matrix for the fit is presented in Figure 3-4. We present accuracies, purities, and completeness scores for each class in Table 3-3.

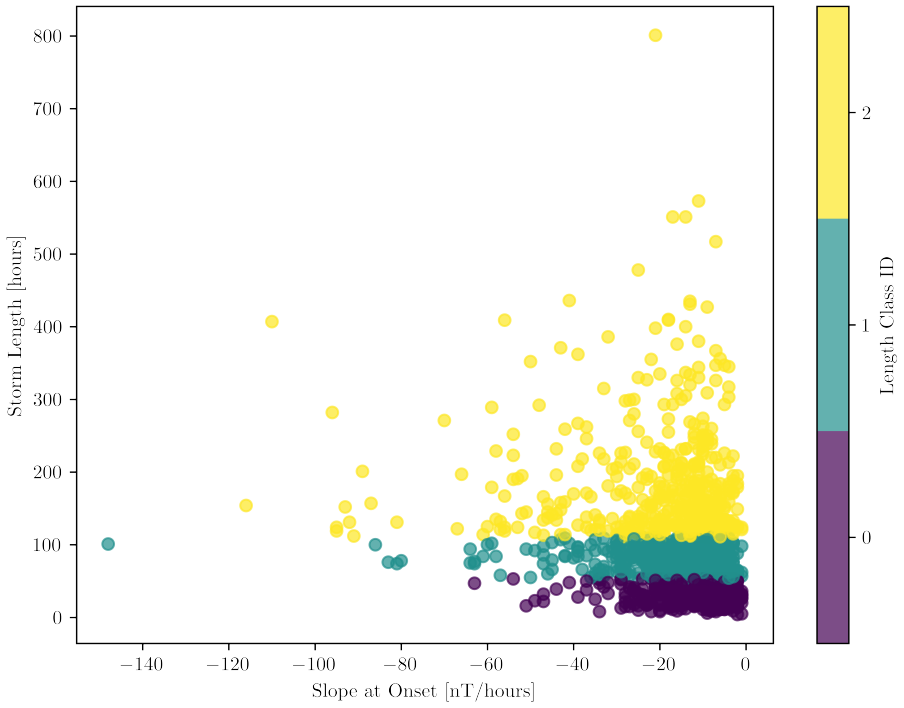



Figure 3-3: Scatterplot of storm length vs. slope at onset.

Table 3-3: Classification metrics for slope at onset.

| Storm Class | Accuracy | Purity | Completeness |
|---|---|---|---|
| Short | 0.56 | 0.43 | 0.58 |
| Medium | 0.51 | 0.28 | 0.23 |
| Long | 0.65 | 0.32 | 0.25 |

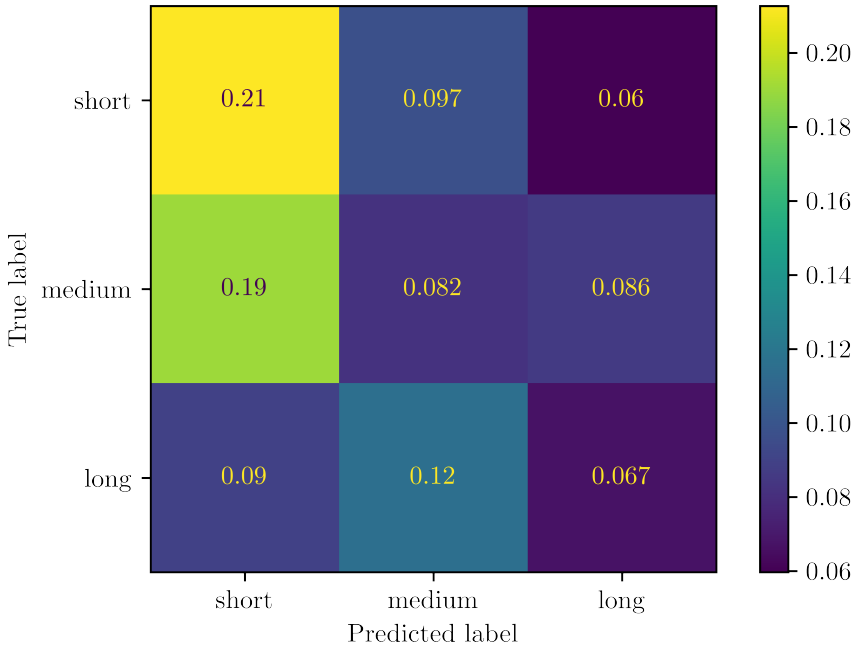



Figure 3-4: Confusion matrix for classification based on slope at onset.

### 3.1.3 Maximum Dst in Day Preceding Storm Onset

In this example, we attempt to predict storm length based on the maximum Dst value in the day leading up to the storm. As demonstrated in Figure 3-5, there is little correlation between this value and storm length. As a result, the classifier is unable to robustly predict the storm length. Overall, the training set is predicted with 40% accuracy while the testing set is predicted with 37% accuracy. A confusion matrix for the fit is presented in Figure 3-6, indicating that medium length storms were identified with mildly higher accuracy while there was no informative prediction on short or long storms. We present accuracies, purities, and completeness scores for each class in Table 3-4.

Table 3-4: Classification metrics for maximum Dst value in day preceding storm.

| Storm Class | Accuracy | Purity | Completeness |
|---|---|---|---|
| Short | 0.56 | 0.35 | 0.32 |
| Medium | 0.60 | 0.39 | 0.45 |
| Long | 0.59 | 0.37 | 0.35 |

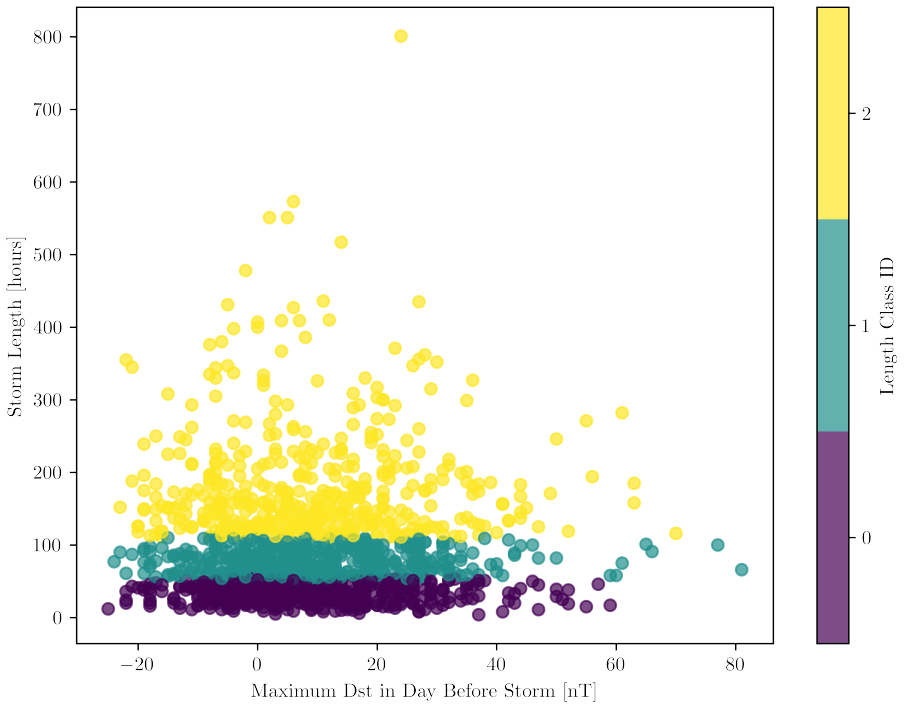



Figure 3-5: Scatterplot of storm length vs. maximum Dst value in day preceding storm.

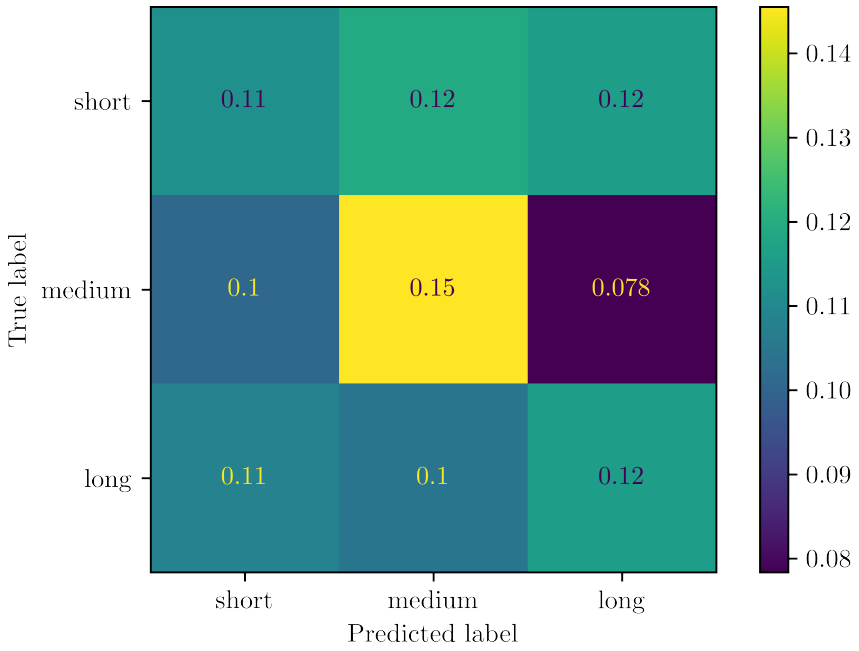



Figure 3-6: Confusion matrix for classification based on maximum Dst value in day preceding storm.

### 3.1.4 Duration of Previous Storm

In this example, we attempt to predict storm length based on the duration of the immediately preceding storm. As demonstrated in Figure 3-7, there is little correlation between this value and storm length. As a result, the classifier is unable to robustly predict the storm length. Overall, the training set is predicted with 40% accuracy while the testing set is predicted with 30% accuracy, indicative of overfitting. A confusion matrix for the fit is presented in Figure 3-6, where the vast majority of short duration storms are misclassified as medium storms. We present accuracies, purities, and completeness scores for each class in Table 3-4.

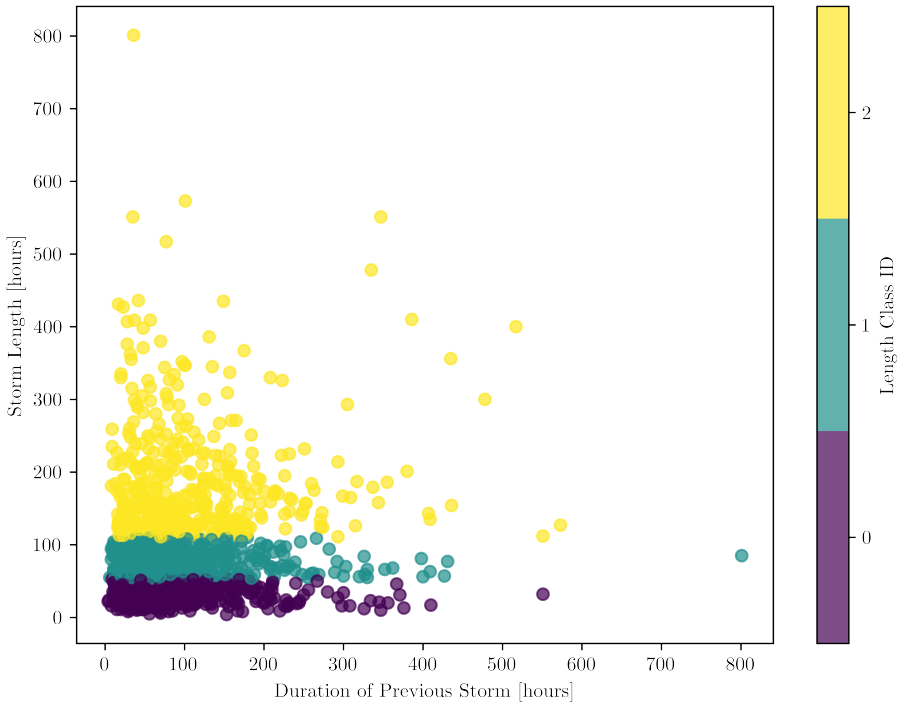



Figure 3-7: Scatterplot of storm length vs. length of immediately previous storm.

Table 3-5: Classification metrics for previous storm length.

| Storm Class | Accuracy | Purity | Completeness |
|---|---|---|---|
| Short | 0.55 | 0.26 | 0.21 |
| Medium | 0.46 | 0.32 | 0.47 |
| Long | 0.59 | 0.32 | 0.22 |

We also attempted to predict storm length based on the length of a storm 27 days (one geomagnetic cycle) prior to a storm. In this case, sample size was too small to produce predictions without severe overfitting.

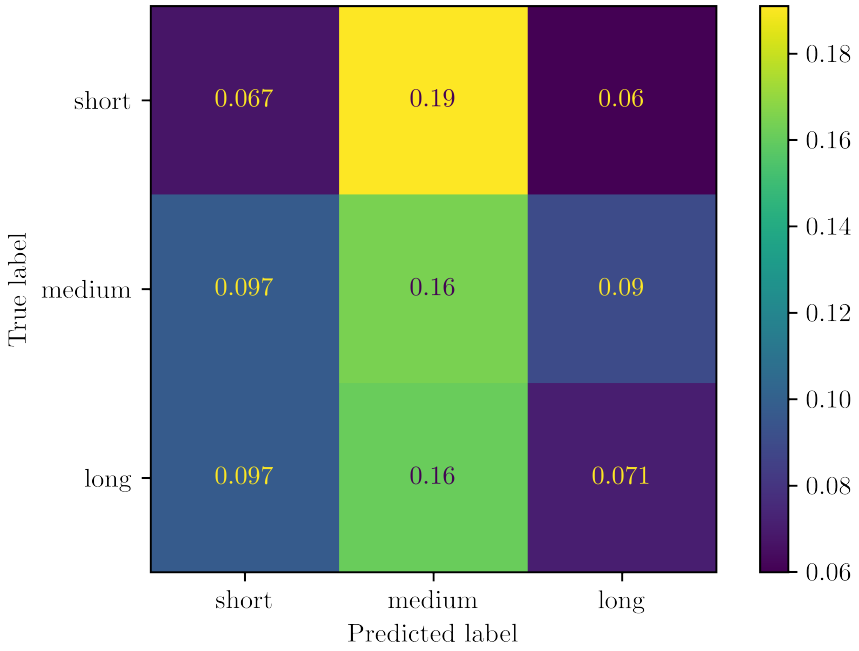



Figure 3-8: Confusion matrix for classification based on length of immediately previous storm.

### 3.1.5 Multiple Dst Values Preceding Storm Onset

We expand our classification scheme to use multiple Dst values leading up to the storm to predict the storm length, including Dst values up to 10 hours before onset. We fit ten separate classifiers, each using different amounts of data for the fit. Initially, we used the Dst value at onset in addition to the value one hour before onset as input. Then, we expanded the classifier to add in the Dst value two hours before onset, then three hours, and so on. We found no significant increase in accuracy when adding in more preceding data; rather, results remained fairly consistent among the ten classifiers. Both testing and training accuracies remained around 42% with deviation of no more than $\sim$5%.

As an example, we present the results for the classifier using hourly Dst values up to six hours before the storm. This classifier has the most consistent accuracies between training and testing sets, indicating little overfitting. A confusion matrix for the fit is presented in Figure 3-9, where the majority of short and medium storms are often classified as short storms, regardless of true duration. We present accuracies, purities, and completeness scores for each class in Table 3-6. The nine other classifiers considered produce categorically similar results.

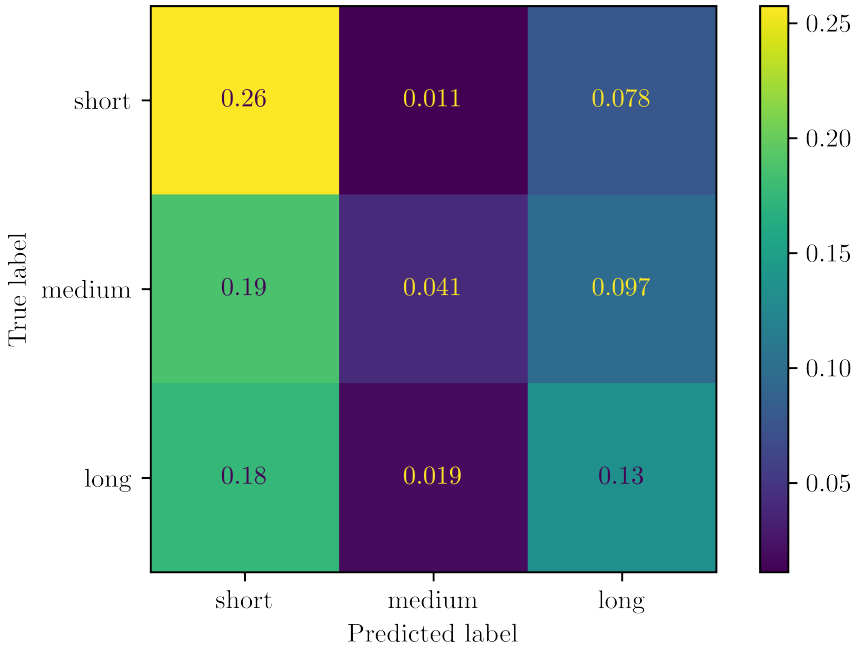



Figure 3-9: Confusion matrix for classification based on hourly Dst data up to six hours before the storm.

Table 3-6: Classification metrics for hourly Dst data up to six hours before the storm.

| Storm Class | Accuracy | Purity | Completeness |
|---|---|---|---|
| Short | 0.55 | 0.42 | 0.74 |
| Medium | 0.69 | 0.58 | 0.13 |
| Long | 0.63 | 0.44 | 0.41 |

### 3.1.6 Combining Parameters

As a final example, we combine every single input parameter used in this section into one classifier, resulting in a classifier with 13 input parameters. We see no signficant increase in accuracy with a larger number of storm parameters. A confusion matrix is presented in Figure 3-10 while metrics for accuracy, completeness, and purity are presented in Table 3-7. This classifier rarely predicts storms as medium storms, regardless of true duration.

Table 3-7: Classification metrics based on 13 combined input parameters.

| Storm Class | Accuracy | Purity | Completeness |
|---|---|---|---|
| Short | 0.57 | 0.44 | 0.73 |
| Medium | 0.67 | 0.48 | 0.16 |
| Long | 0.65 | 0.45 | 0.43 |

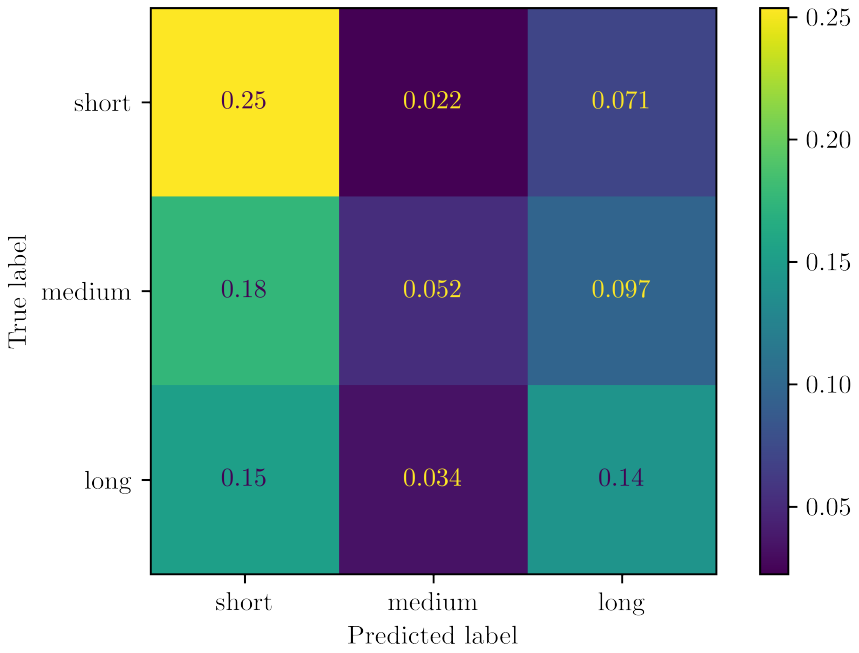



Figure 3-10: Confusion matrix for classification based on 13 input parameters.

### 3.1.7 Classification Summary

None of the classifiers presented in this section recover storm duration with reasonable accuracy. These classifiers have no meaningful utility for prediction storm length. Although not presented here, we also attempted to predict the minimum Dst value during a geomagnetic storm from similar input parameters. This resulted in even less accurate predictions. We also note that mild overfitting remains present in several of the classifiers presented in this section, despite efforts to mitigate overfitting. Random forest classifiers are naturally suited to avoid overfitting, but overfitting is difficult to avoid when there are no clear trends to exploit to infer output parameters. Lastly, we note that all the classification metrics provided here rely on Dst data alone. Classification may improve by considering additional properties, such as electric field or pressure, in addition to Dst.

## 3.2 Regression

We repeated the classification analysis with a random forest regressor, which seeks to infer the numerical duration of a storm rather than the category of storm duration (short, medium, or long). All regressors are trained on an identical set of input parameters to those in Section 3.1. Similarly to the classification results, to regressor is able to predict storm duration or maximum depth of a geomagnetic storm with any reasonable accuracy. In addition to random forest regression, we tested several other regressors, including Simple Ordinary Least Squares Linear Regression, Kernel Ridge Regression, and Support Vector Regression. None of these regressors produced reasonable predictions.

As an example, Figure 3-11 shows predicted Dst based on ten Dst values leading up to storm

onset. We use a random forest regressor, with hyperparameters tuned through three-fold cross-validation. The results presented are for the testing set. The regressor is unable to infer storm duration.

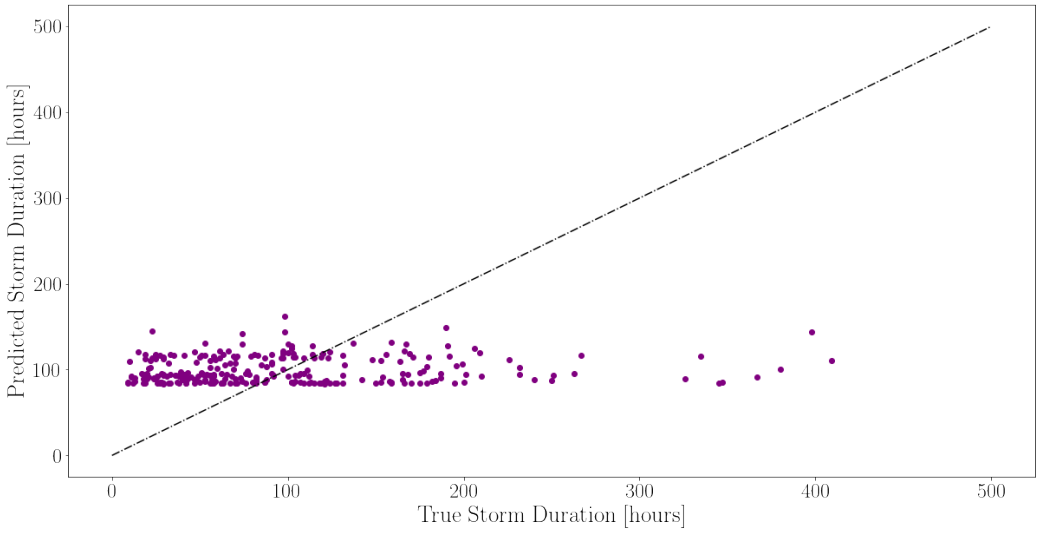



Figure 3-11: Predicted storm duration using random forest regression.

# 4 Time Series Forecasting with Gaussian Process Regression

In this section, we attempt to use Gaussian Process Regressors (GPRs) to extrapolate geomagnetic storm behavior based on previous Dst inputs. GPRs (also known as kriging) interpolate data by modeling a stochastic process where input data convolved with some functional form (kernel). For a complete description of GPR, see Reference [9]. **In general, GPRs are rarely competitive tools for forecasting (extrapolation) and are better suited for interpolation.**

We train a GPR on a lengthy time series (∼7,000 hours) of OMNI data in an attempt to predict future geomagnetic storm behavior. We downsample the Dst data by 20% to decrease the time required to train a GPR. Then, we train the GPR on the 20% of data, then use the GPR to predict the next 100 hours of Dst data. We evaluate the GPRs performance by computing the mean-squared error (MSE, see Equation 4) of all 100 predicted Dst values to the true time series. Using this MSE, we performed kernel and hyperparamter optimization, by testing numerous kernels and GPR parameters to identify the lowest mean-squared error on the predicted data using three-fold cross-validation.

The GPR with the lowest MSE is presented in Figure 4-1. In this figure, the true Dst data are presented in blue, with multiple geomagnetic storms occuring in this 6455 hour period. Red points indicate the downsampled data used for training, while the black curve with gray interval indicates the GPR prediction with associated 90% credible level uncertainty. This GPR is fit with the following kernel: RationalQuadratic(alpha=0.103, length_scale=23) + $0.407^2$ * RBF(length_scale=4.8) + ExpSineSquared(length_scale=0.891, periodicity=647). Figure 4-2 offers a zoomed in view of the same GPR. In this example, the GPR was unable to predict the nearby presence of a storm.

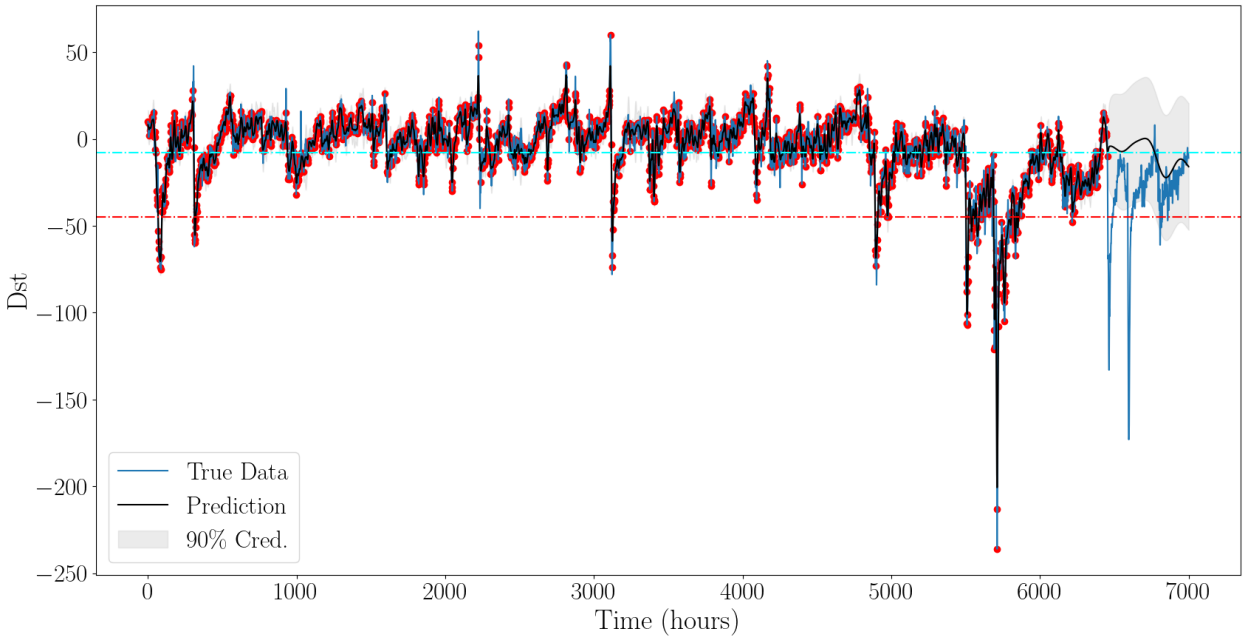



Figure 4-1: GPR trained on 6455 hours of Dst data.

We attempt to improve the GPR fit by including all 100 hours of data prior to the point where we begin extrapolation, instead of randomly downselecting the data. The result of this fit is presented in Figure 4-3. There is no significant improvement in the GPR extrapolation. In Figure 4-4, we

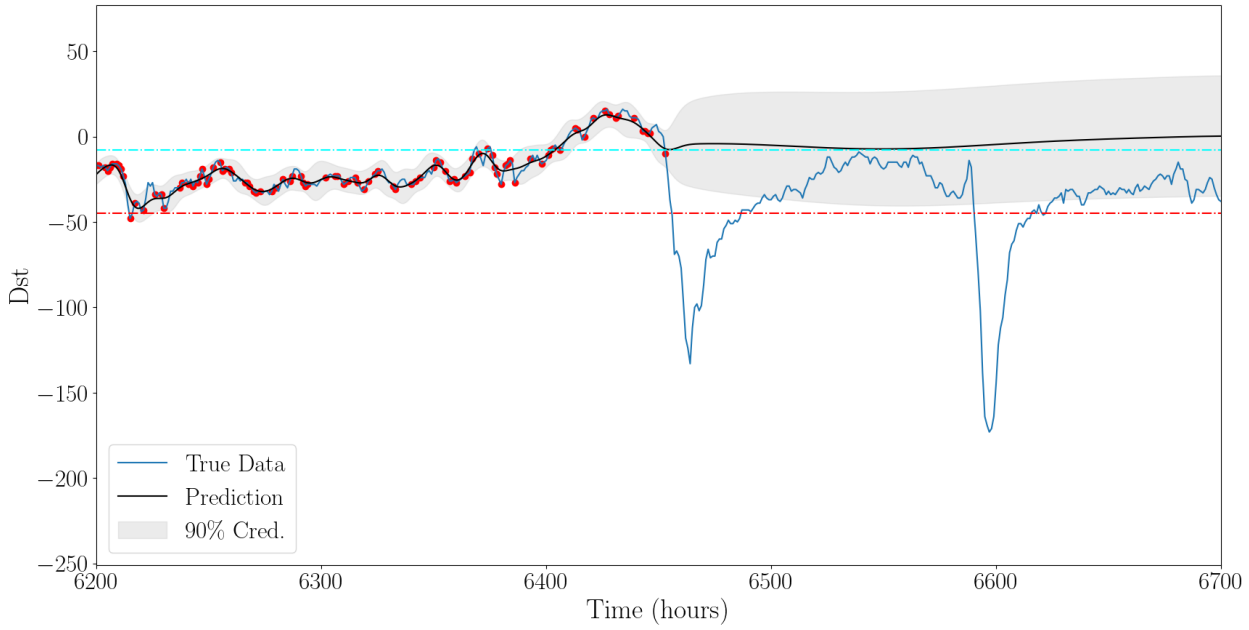



Figure 4-2: Zoomed in image of GPR presented in Figure 4-1

train on one additional hour of Dst data, which does not result in a storm. We add another hour in Figure 4-4 and again in Figure 4-5. Neither of these examples result in a geomagnetic storm.

None of our GPR attempts are able to predict geomagnetic storms. This can be attributed to three possibilities: 1) GPRs are not appropriate extrapolation tools; 2) Dst data offers no predictive power for inferring future Dst; or 3) the GPR models presented in this section are not well trained or exhibit overfitting. The latter point is supported as all GPR extrapolations revert to a mean Dst of zero, which may not be appropriate for the OMNI Dst dataset.

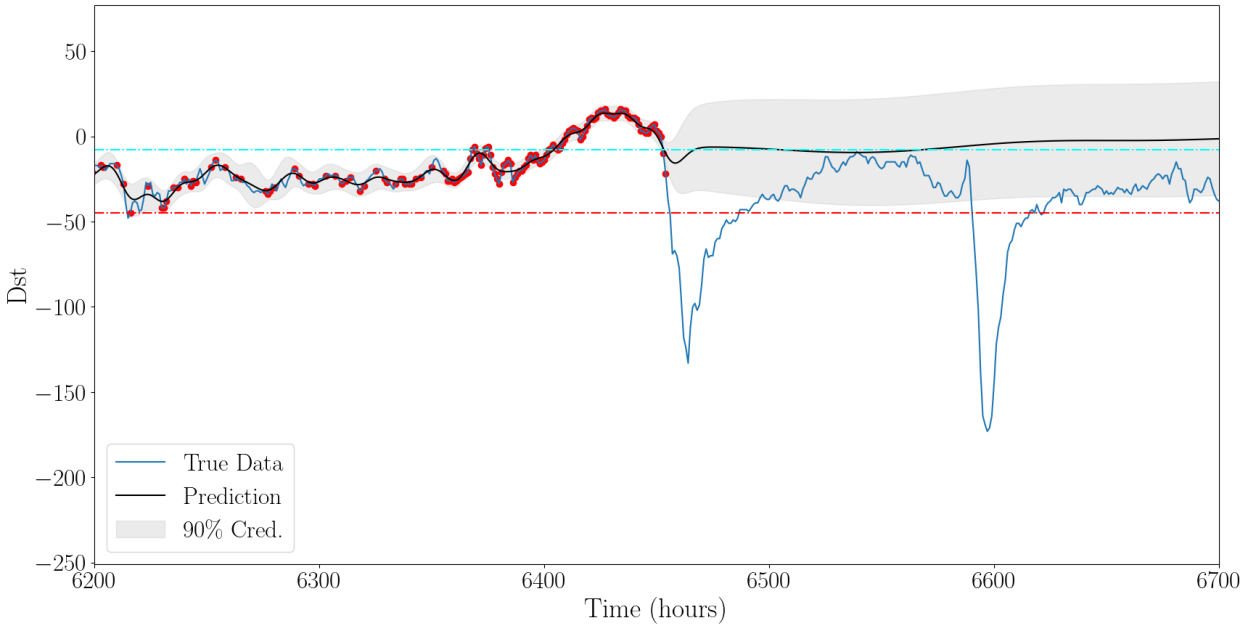



Figure 4-3: GPR trained on 6455 hours of Dst data, with no downsampling in final 100 hours.

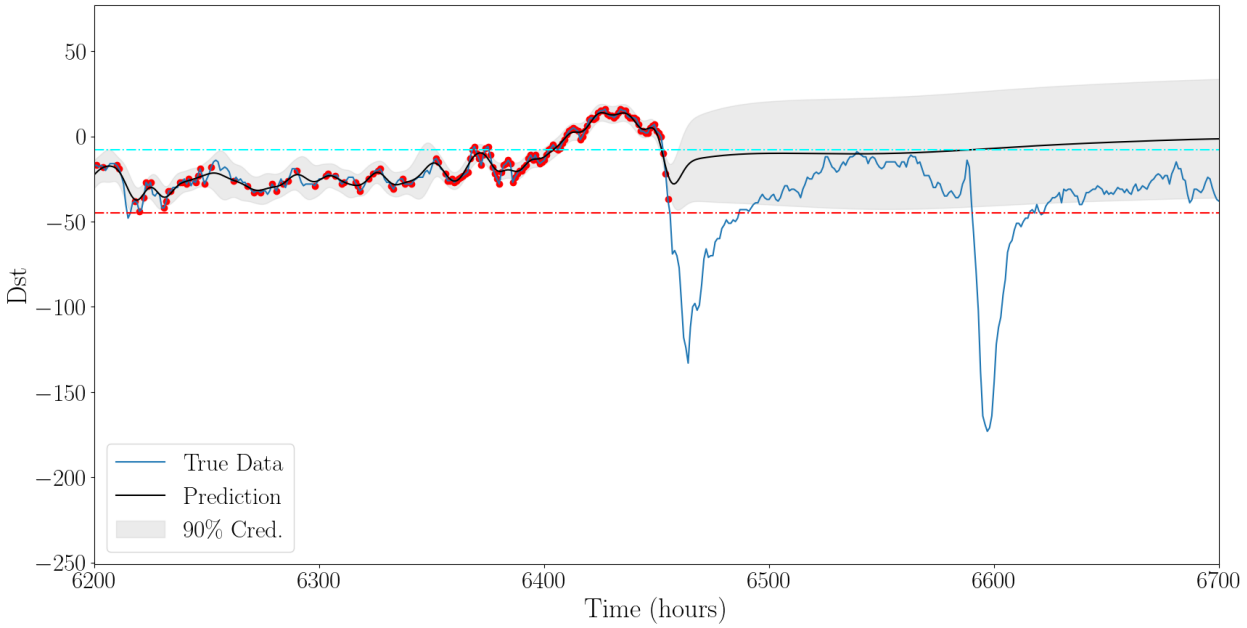



Figure 4-4: GPR trained on 6456 hours of Dst data.

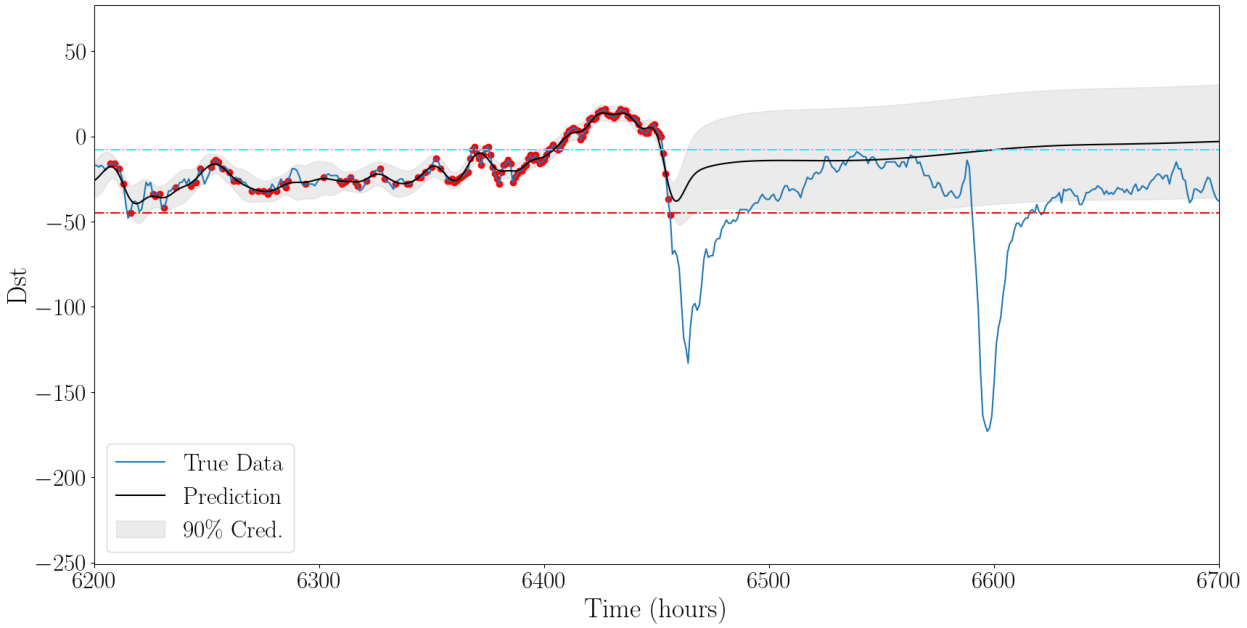



Figure 4-5: GPR trained on 6457 hours of Dst data.

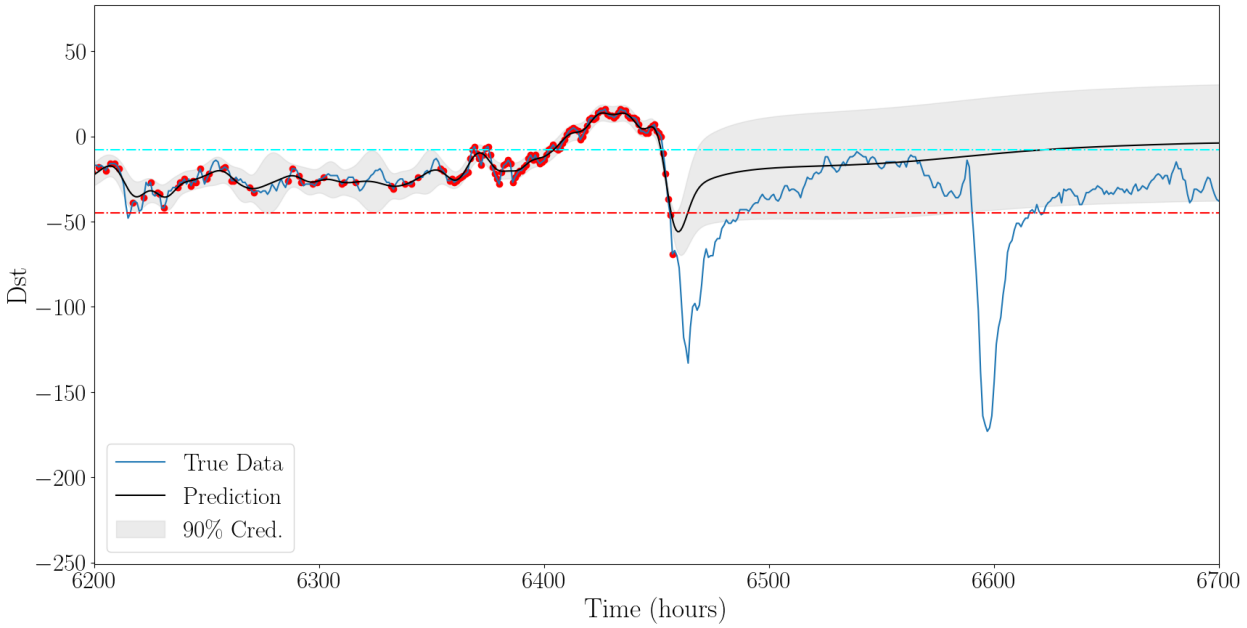



Figure 4-6: GPR trained on 6458 hours of Dst data.

# 5 Time Series Forecasting with Sliding Windows

In this section, we use time slides coupled with machine learning regression to forecast Dst time series. We use a random forest regressor, with hyperparameters tuned with five-fold cross-validation. In our initial example, we use the current Dst value to predict the Dst value one hour in the future, such that input parameters correspond to $\mathrm{Dst}(t)$ and output parameters are $\mathrm{Dst}(t+1)$. Figure 5-1 indicates the correlation between $\mathrm{Dst}(t+1)$ and $\mathrm{Dst}(t)$ for a year of OMNI data. There is a clear correlation, which can easily be exploited to infer the next timestep. Although not necessary for a simple linear relationship, we train a random forest classifier to predect Dst one hour in advance given the current Dst value. The predicted and true values are presented in Figure 5-2, where the light gray curve indicates the true data while the blue curve with black points indicates the predicted data. The red point indicates the point after which the first prediction is made. In this example, the regressor is trained on over seven months of Dst data in 2010. We measure the accuracy of the fit with a mean-squared error (MSE), listed in the title of the bottom panel of Figure 5-2. MSE is computed by

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=0}^{n-1}\left(y_i - \hat{y}_i\right)^2, \tag{4}$$

where $n$ is the total number of predicted samples, $\hat{y}_i$ is the predicted value for one sample, and $y_i$ is the true value for the corresponding sample. Smaller MSE values indicate a more accurate fit, although overfitting may still be possible with small MSE values.

While the prediction in Figure 5-2 corresponds to a comparatively low MSE, the fit is not particularly robust. In this example, the regressor assumes that $\mathrm{Dst}(t) \approx \mathrm{Dst}(t+1)$ with little variation. This is not a strong relationship for inferring storm behavior beyond one hour.

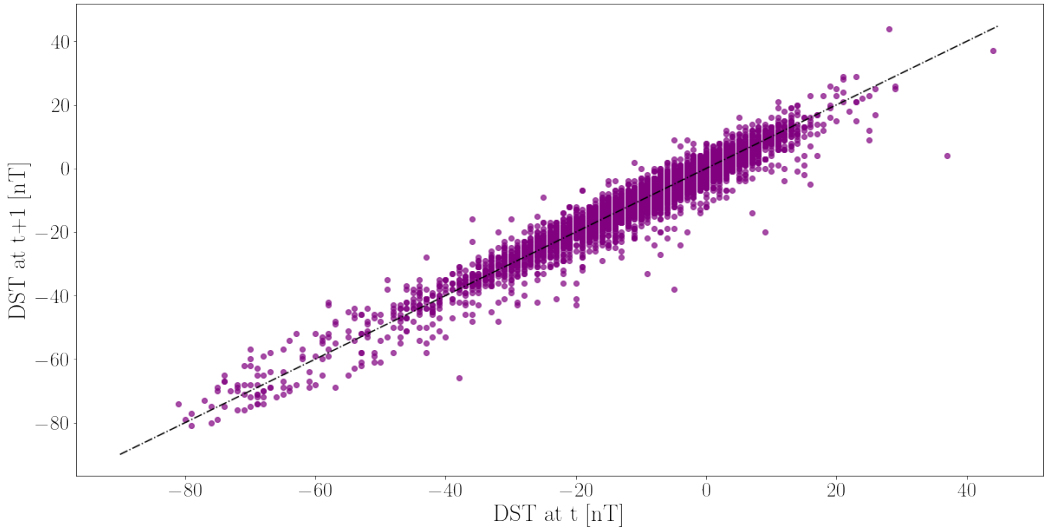



Figure 5-1: Correlation between Dst and Dst one hour ahead.

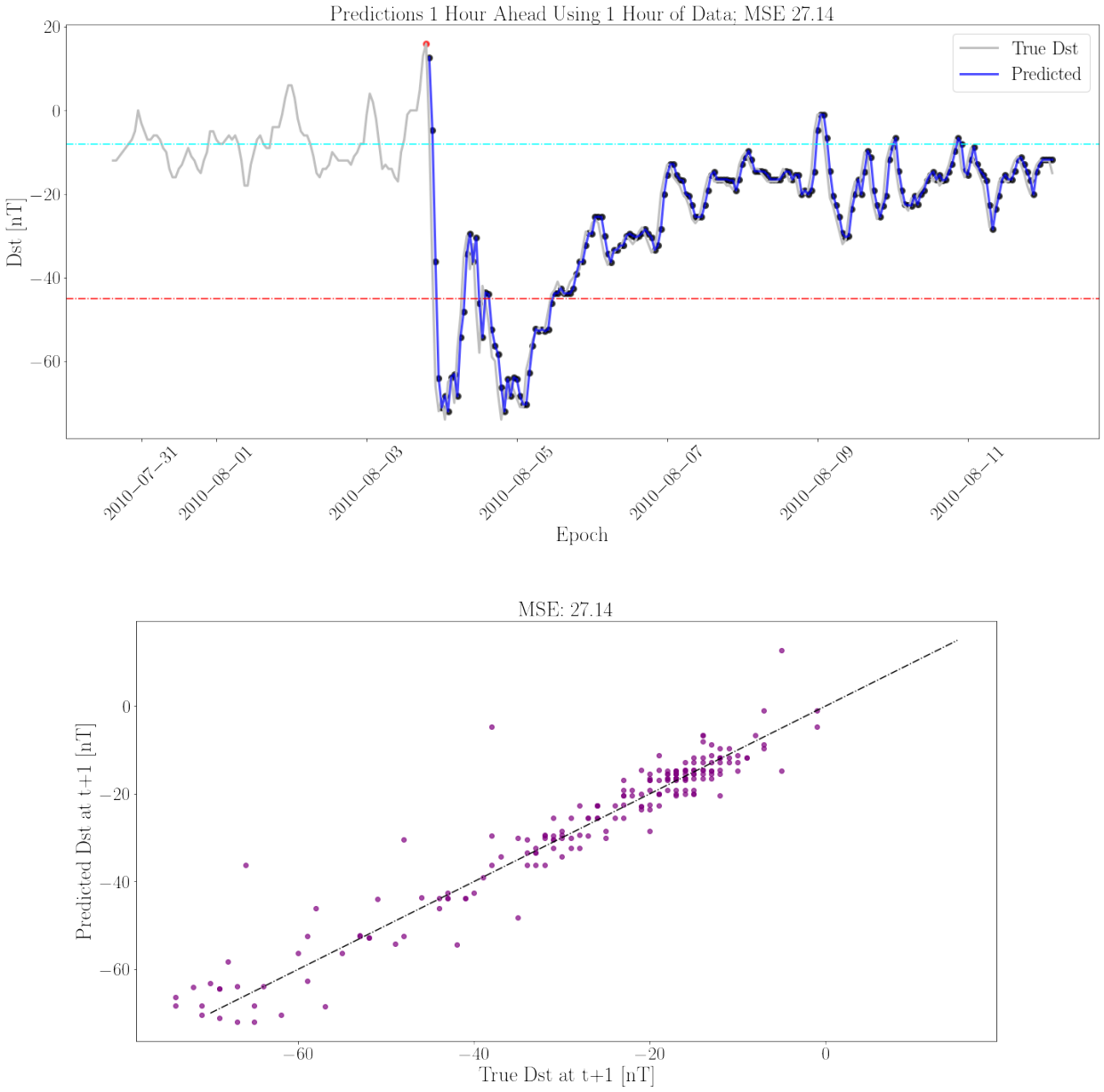



Figure 5-2: Predicting Dst one hour in advance.

## 5.1 Predicting Later Times

Instead of predicting Dst one hour in advance, we can predict Dst at some arbitrary $N$ number of hours in advance. In all examples in this section, the regressor is trained on over seven months of Dst data in 2010.

Figure 5-3 displays the correlation between $\mathrm{Dst}(t+2)$ and $\mathrm{Dst}(t)$. Again, a linear correlation is clear, promoting estimates of Dst. The predicted Dst is displayed in Figure 5-4, where the regressor tends to predict that $\mathrm{Dst}(t) \approx \mathrm{Dst}(t+2)$. Unlike Figure 5-2, this assumption begins to break down prediction accuracy, resulting in a much larger MSE.

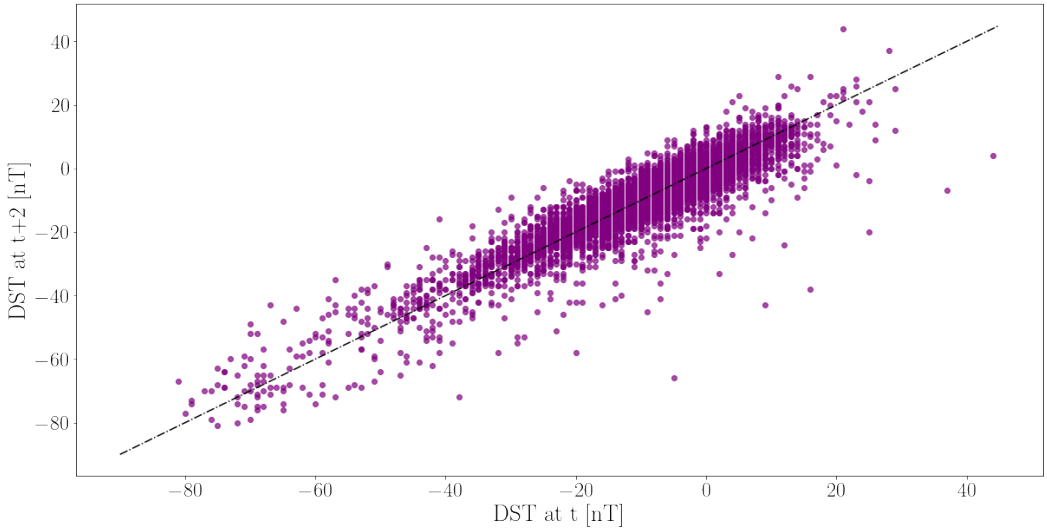



Figure 5-3: Correlation between Dst and Dst two hours ahead.

Using an identical method, we infer Dst for six, 12, and 24 hours in advance. Correlations in Dst for these values are presented in Figure 5-5. Outlier Dst values are uncorrelated with their preceding Dst values for significant time differences. Machine learning predictions on Dst are presented in Figures 5-6 through 5-8, where Dst is poorly predicted for all examples.

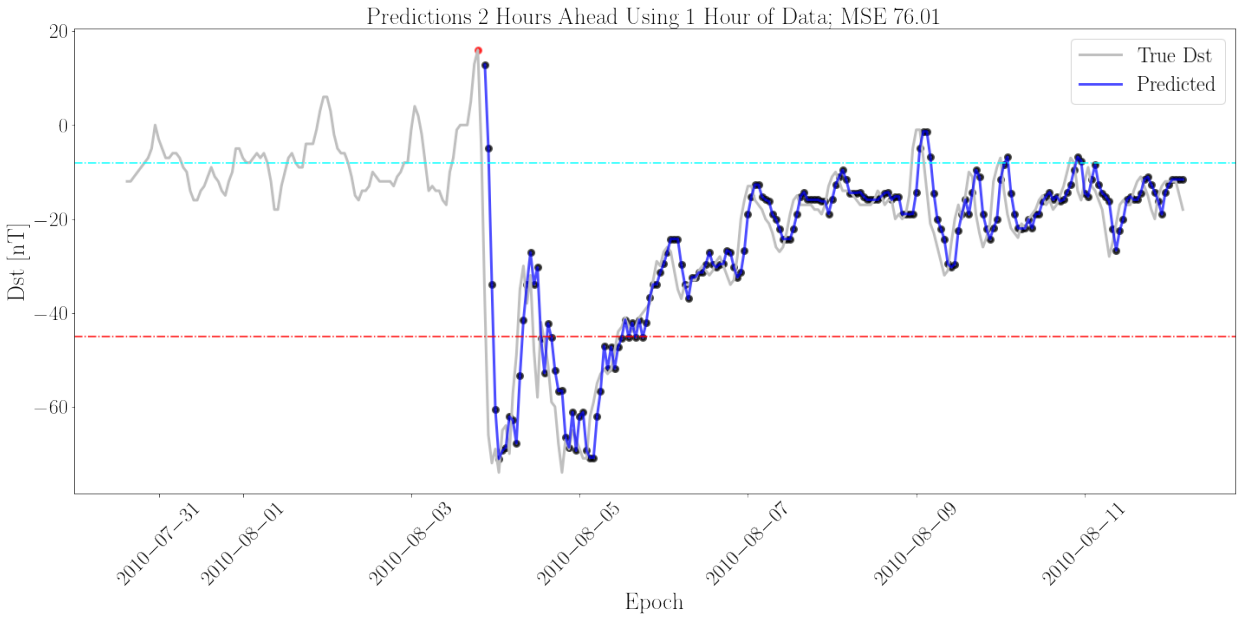



Figure 5-4: Predicting Dst two hours in advance.

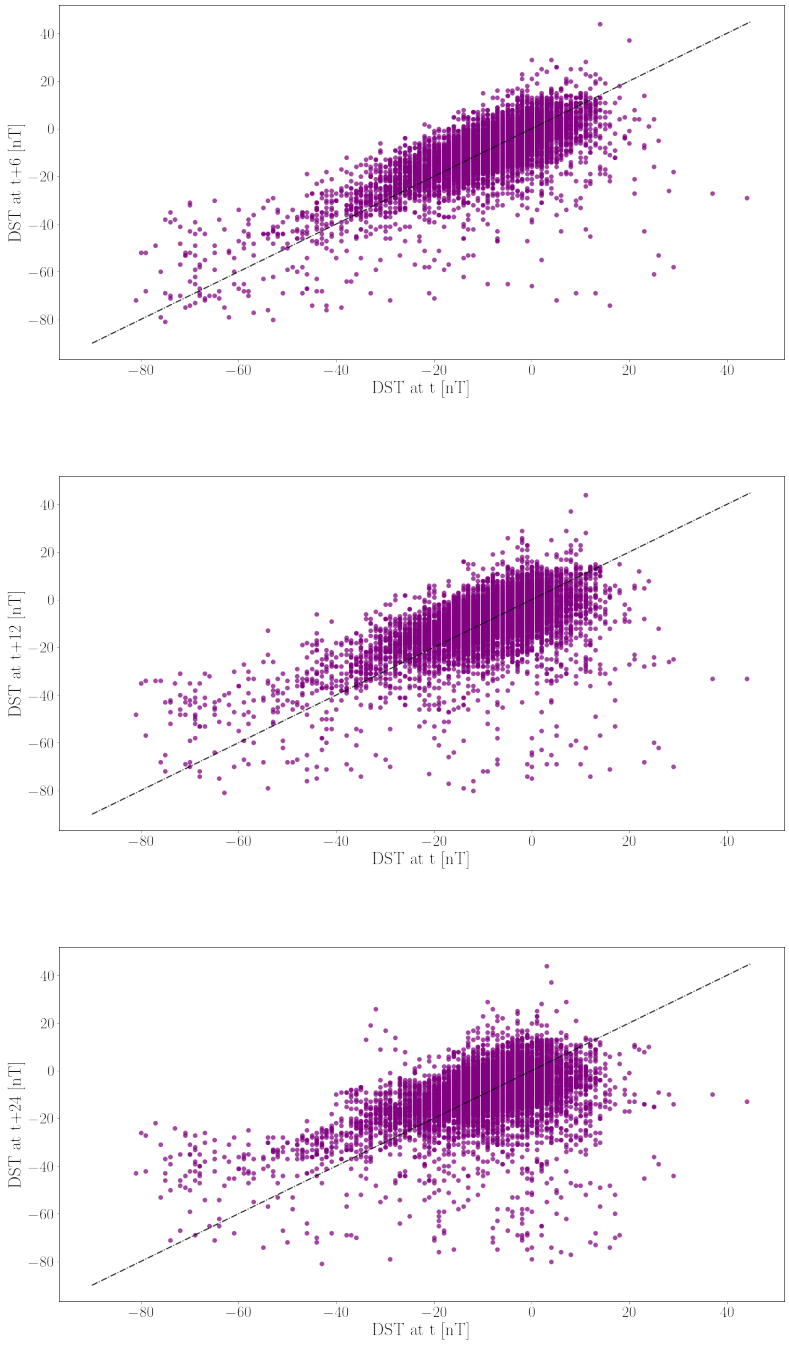



Figure 5-5: Correlations in Dst for increments of six, 12, and 24 hours.

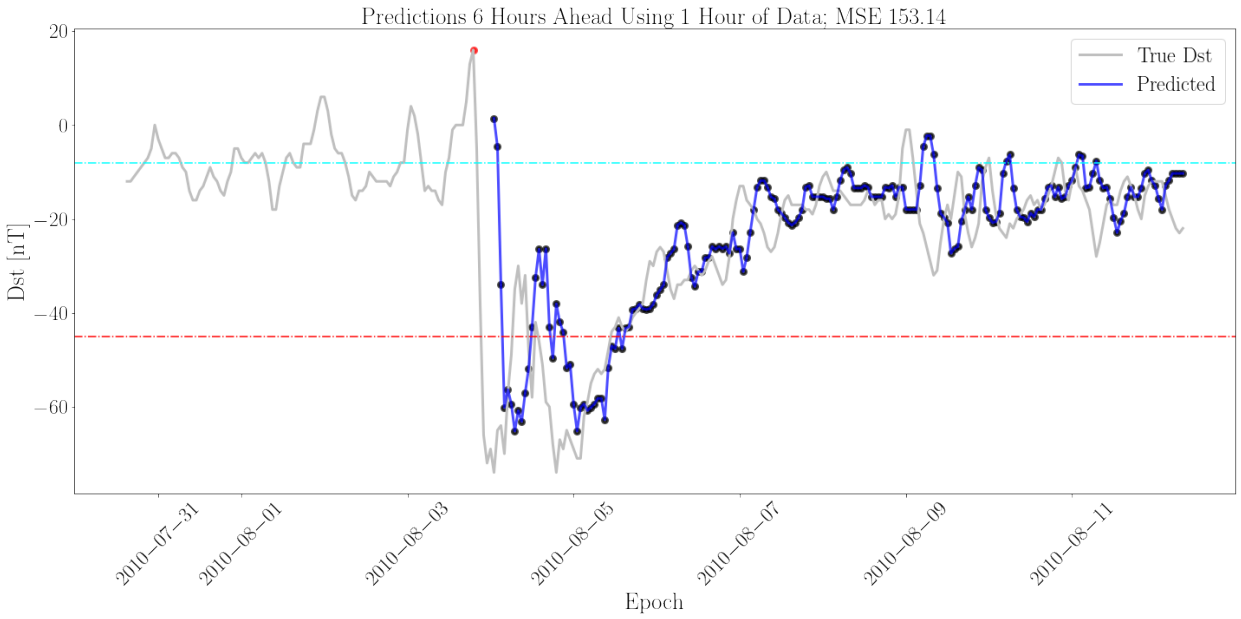



Figure 5-6: Predicting Dst six hours in advance.

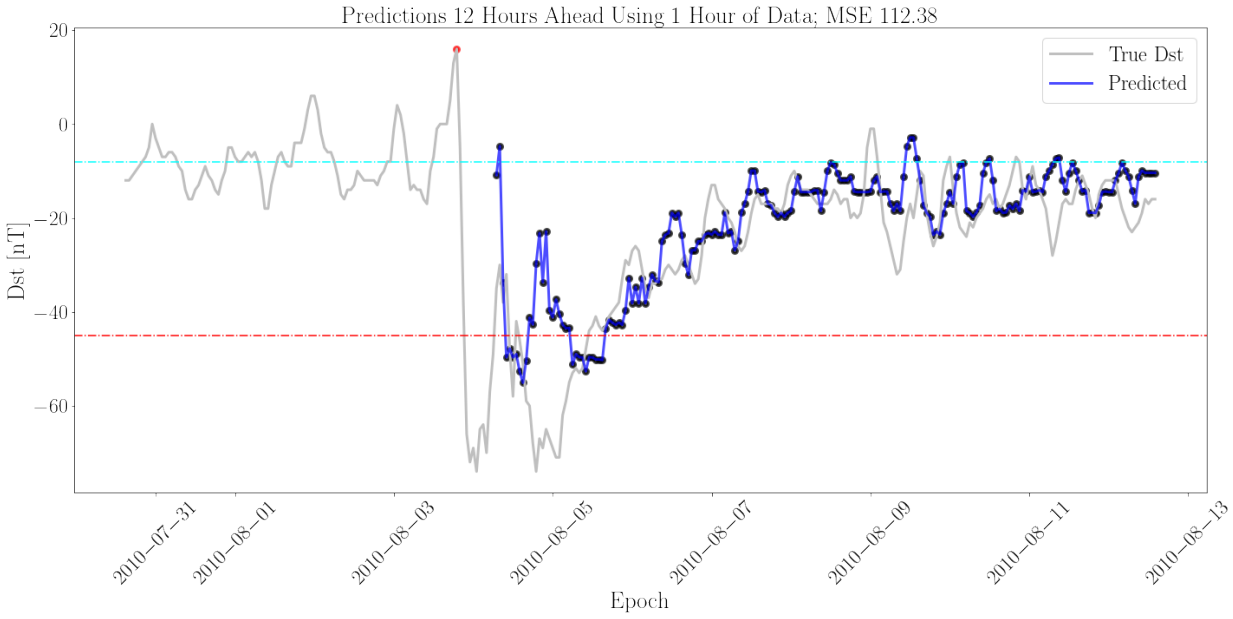



Figure 5-7: Predicting Dst 12 hours in advance.

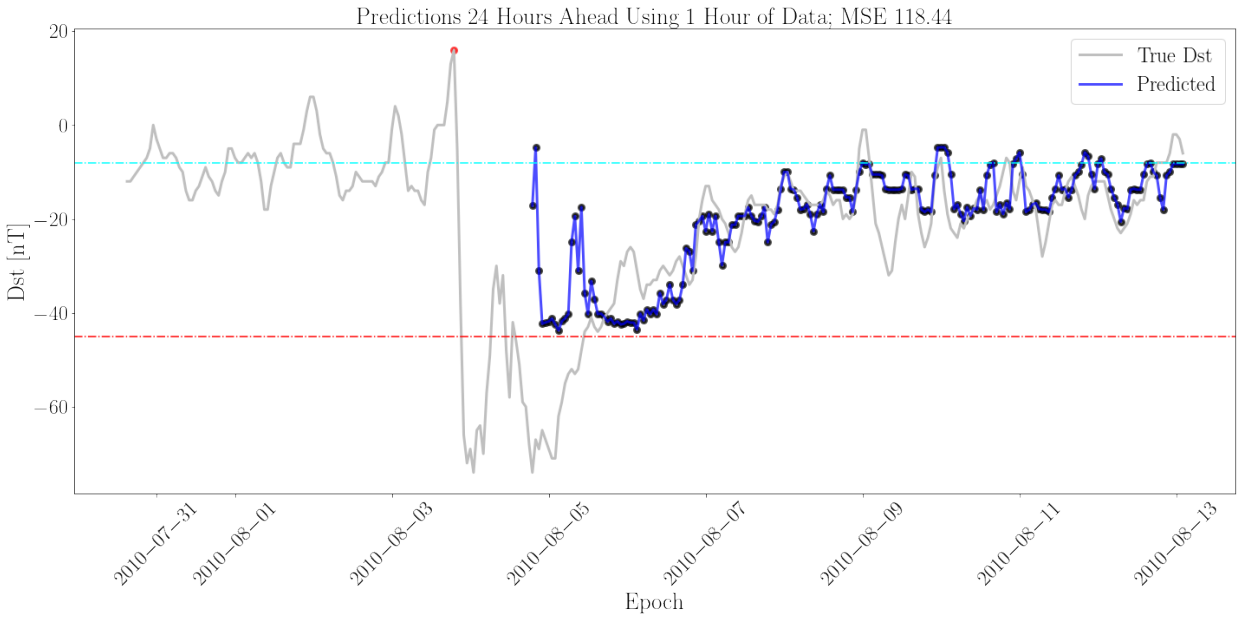



Figure 5-8: Predicting Dst 24 hours in advance.

## 5.2 Leveraging Pressure and Electric Field

We can leverage additional time-varying data along with Dst to predict storms. In particular, pressure and electric field are anticipated to be key components in Dst time series [5]. Figures 5-9 and 5-10 represent the correlation between each parameter at time $t$ with $\text{Dst}(t+1)$. While no correlations are immediately clear (except perhaps a mild anti-correlation with electric field), these parameters can still be used in tandem to enhance time series forecasts.

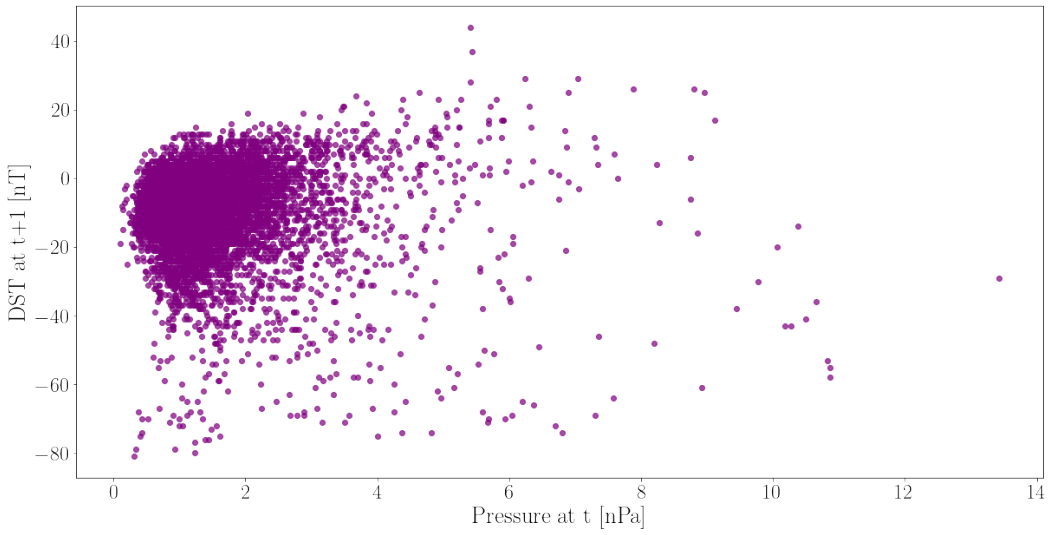



Figure 5-9: Correlation between pressure and Dst.

We repeat all previous regression examples using input data of both Dst and electric field at time $t$ to infer Dst at a future time. In Figure 5-11, we present an estimate on Dst, given the Dst and electric field values one hour prior. The addition of electric field breaks the regressor's tendency to assign $\text{Dst}(t) \approx \text{Dst}(t+1)$ and improves prediction accuracy: compared to Figure 5-2, the MSE has decreased from 27.14 to 23.55.

In Figure 5-12, we include pressure one hour prior to prediction, in addition to Dst and electric field data, resulting in a reduction of MSE to 18.60. We repeat this analysis for predictions more than one hour in advance, and uniformly see improvement in prediction accuracy when including electric field and/or pressure data. These improved predictions are presented in Figures 5-13 through 5-16.

We note that for some examples, electric field and Dst produce a better prediction than electric field, Dst, and pressure together. In general, it seems that electric field enhances predictions between one and three hours in advance, while pressure significantly increases accuracy in predictions over four hours in advance.

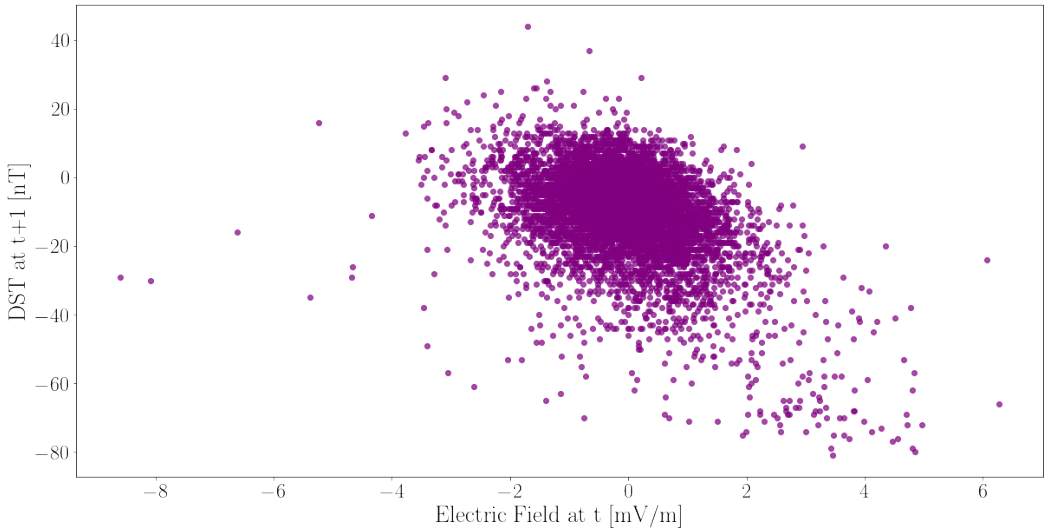



Figure 5-10: Correlation between electric field and Dst.

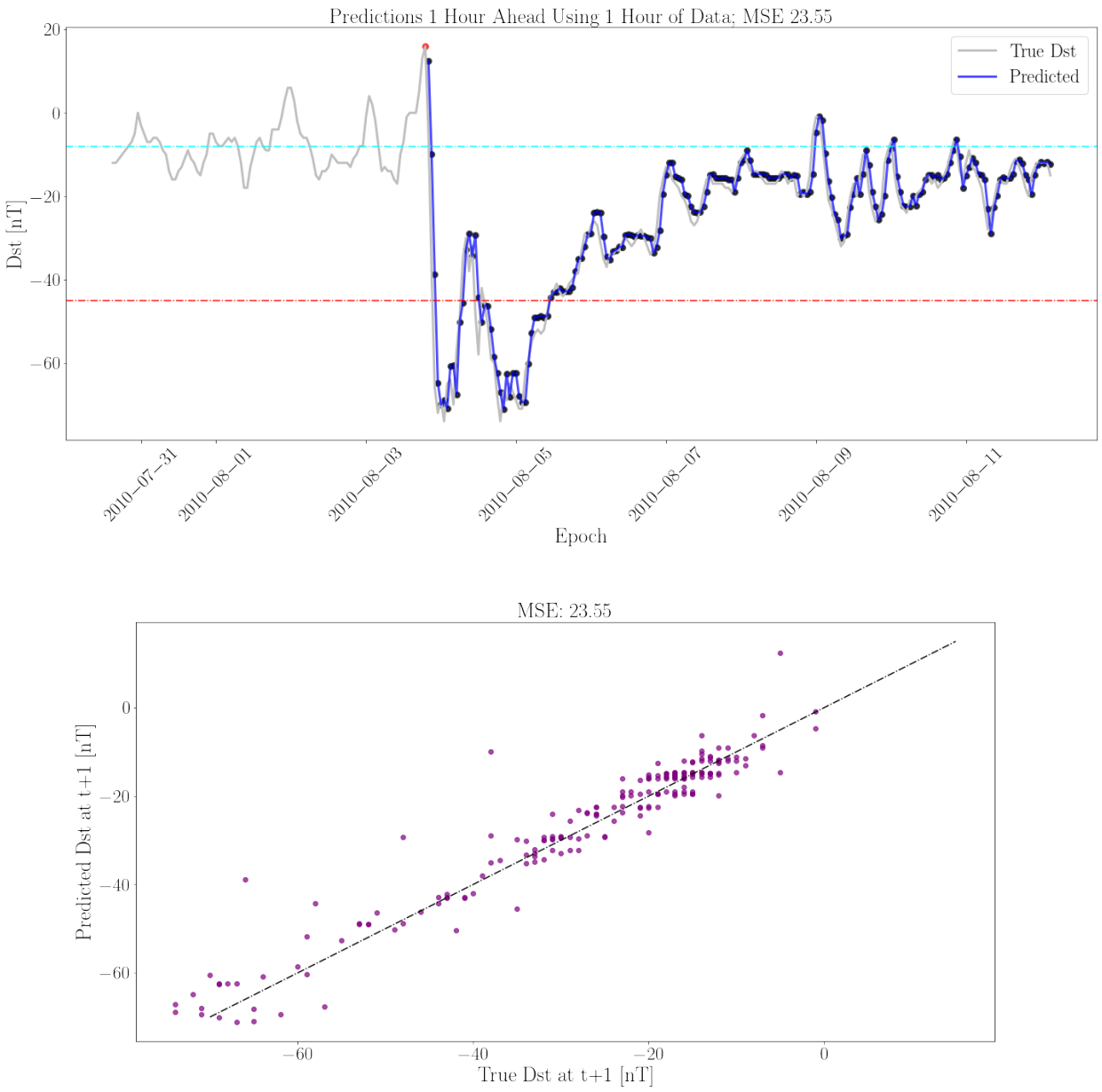



Figure 5-11: Predicting Dst one hour in advance using both Dst and electric field input data.

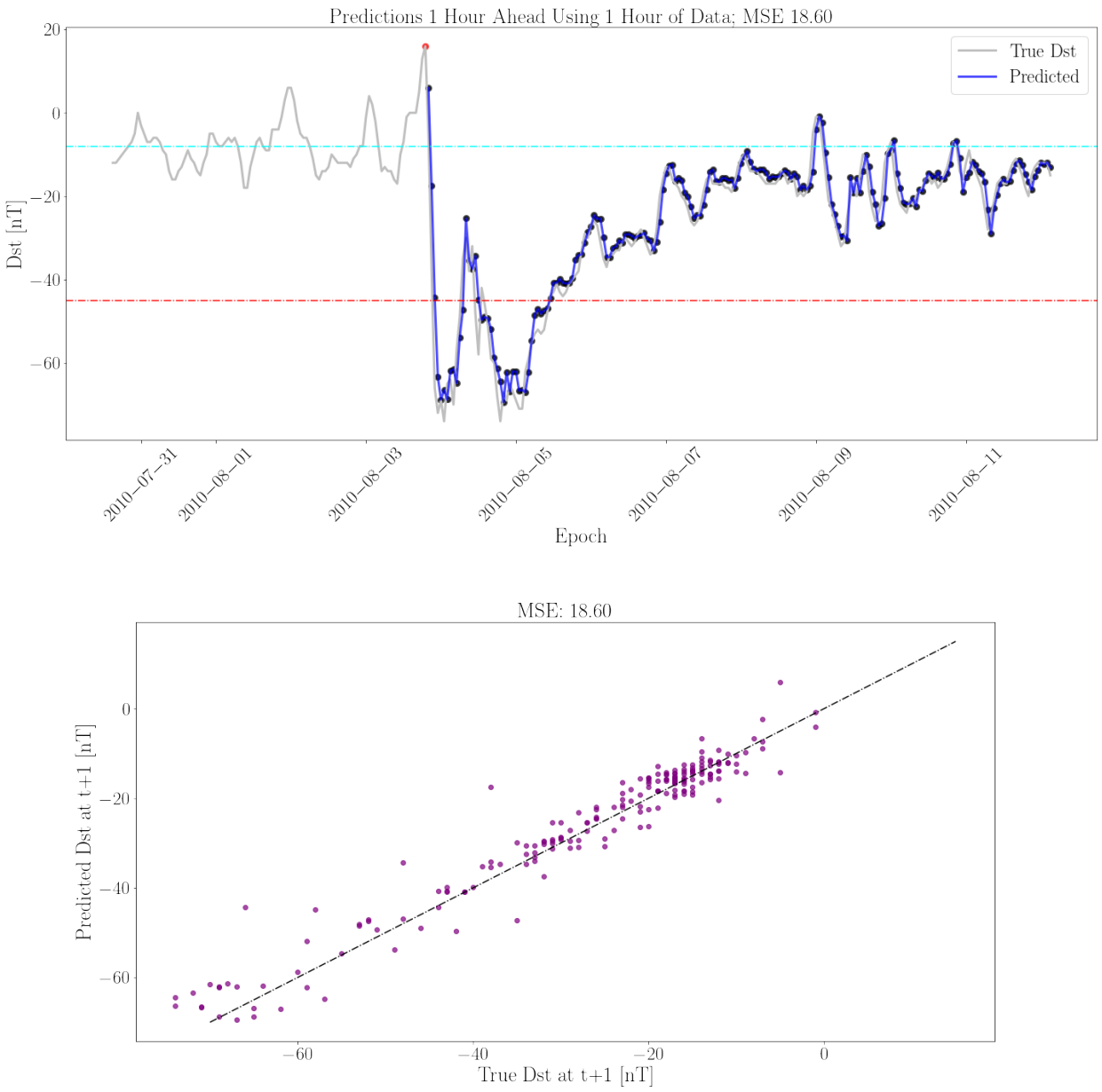



Figure 5-12: Predicting Dst one hour in advance using Dst, electric field, and pressure input data.

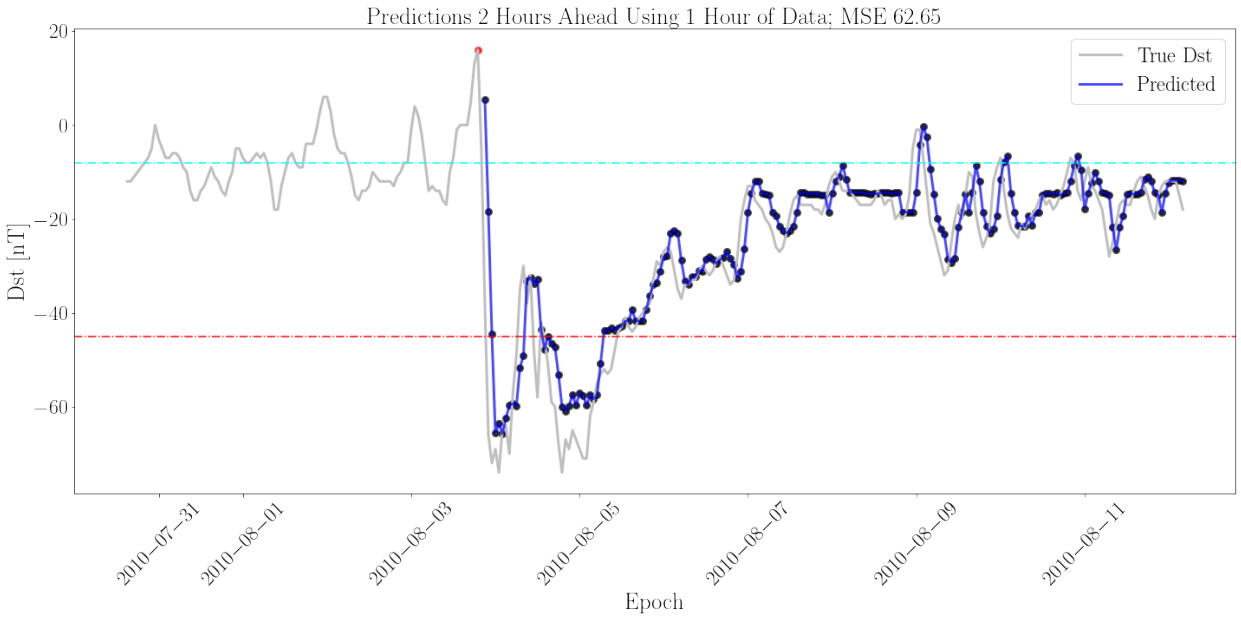



Figure 5-13: Predicting Dst two hours in advance using Dst, electric field, and pressure input data.

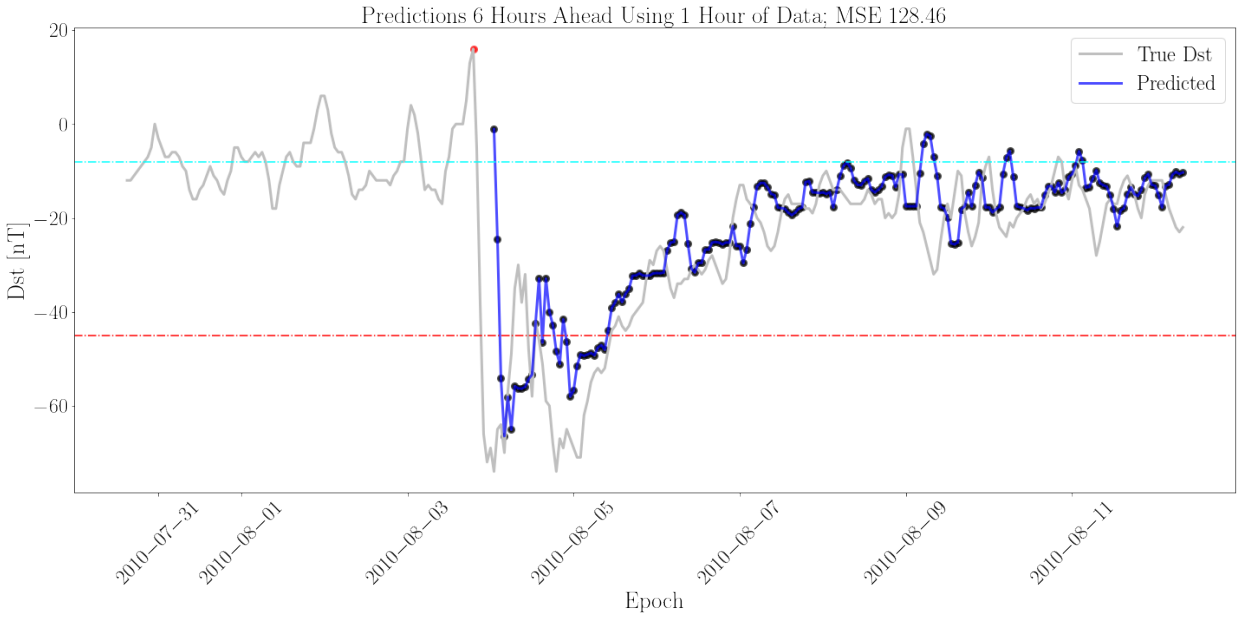



Figure 5-14: Predicting Dst six hours in advance using Dst, electric field, and pressure input data.

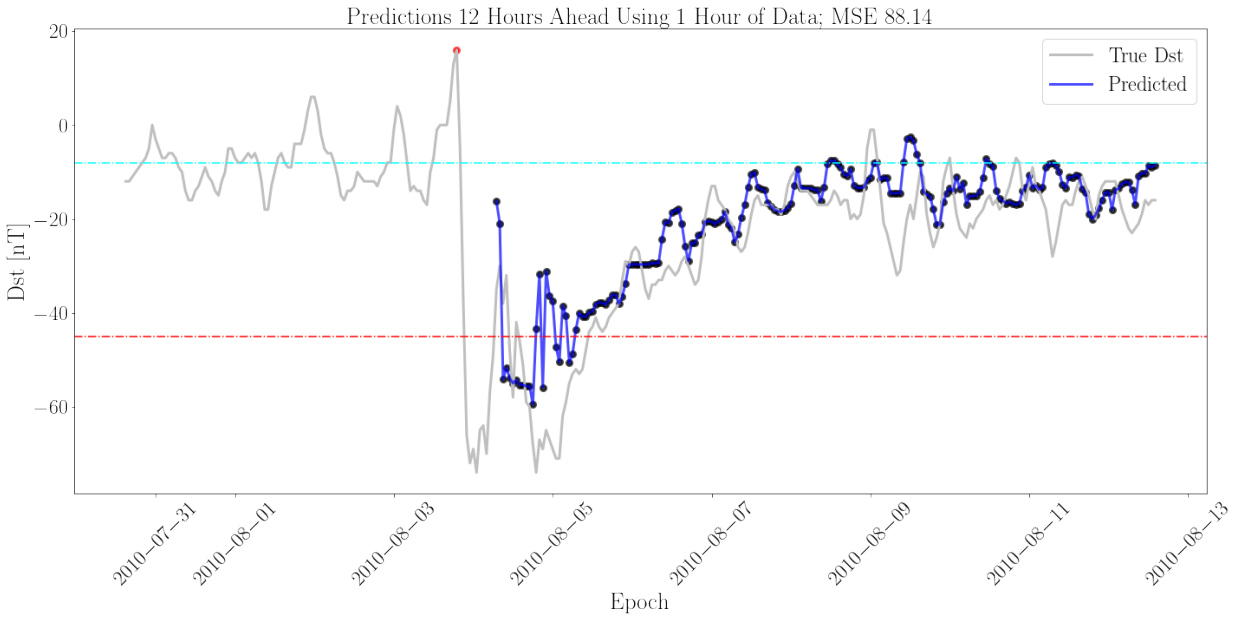



Figure 5-15: Predicting Dst 12 hours in advance using Dst, electric field, and pressure input data.

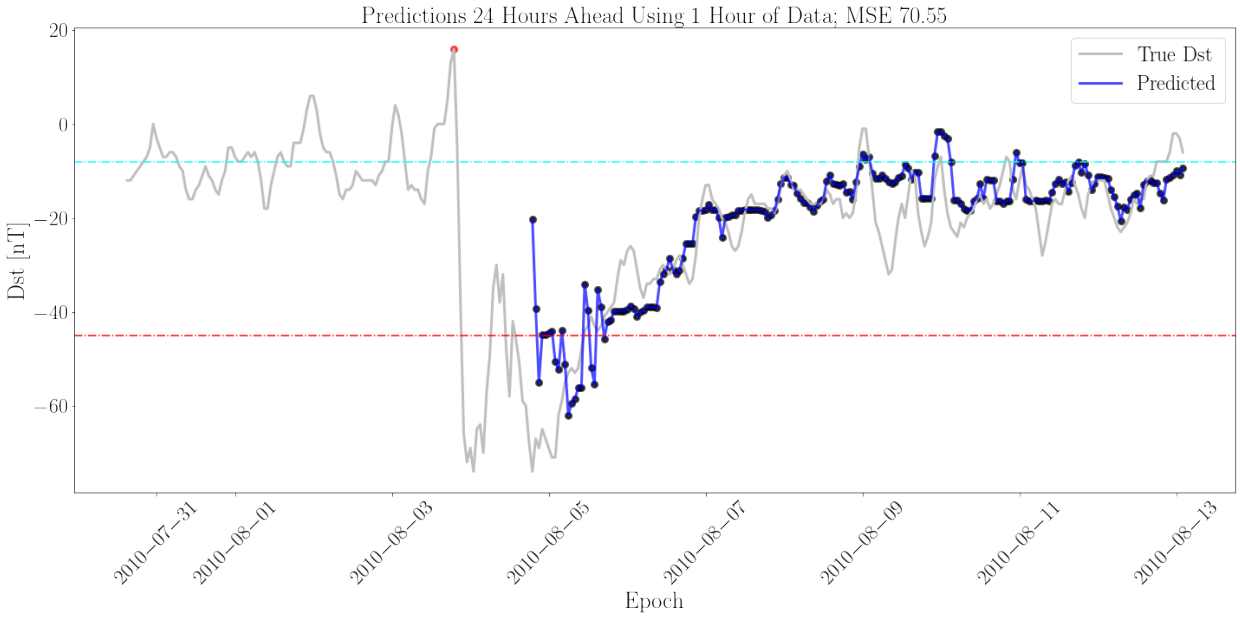



Figure 5-16: Predicting Dst 24 hours in advance using Dst, electric field, and pressure input data.

## 5.3 Using Additional Preceding Data

We can further expand the input data to the random forest regressor by including Dst values at multiple timesteps leading up to the value we are attempting to predict. In Figure 5-17, we present predictions for Dst using Dst, electric field, and pressure values at both one and two hours prior to the prediction. For the example of predicting Dst one hour in advance, an additional hour of data does not enhance the prediction.[2] We expand these examples using multiple hours of previous data, which are not displayed in this document. We find that the best prediction (lowest MSE) for predicting Dst one hour in advance results from using ten hours of previous data, although improvements to MSE are within the margin of error. For most examples, predictions continue to improve as additional hours of previous data are used for predictions, up to 24 hours of previous data.

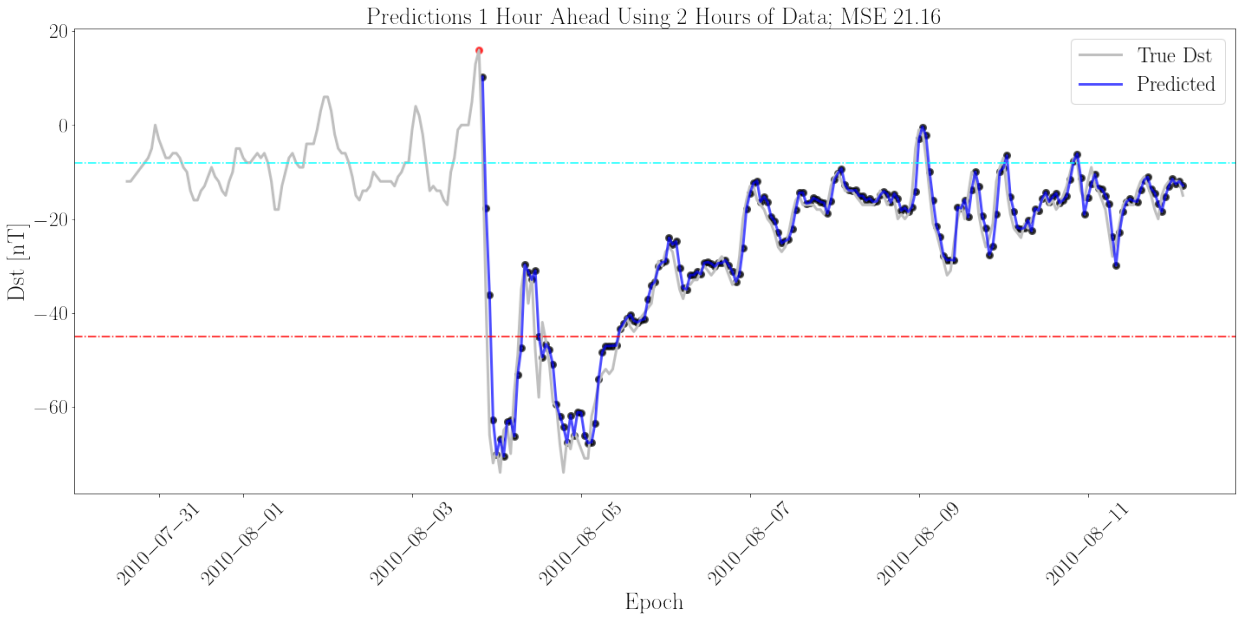



Figure 5-17: Predicting Dst one hour in advance using two hours of previous Dst, electric field, and pressure data as input.

## 5.4 Forecasting Dst

We combine elements of all the previous estimates to produce a tool to forecast Dst up to 24 hours in advance. The tool shows promise and occasionally forecasts Dst to high accuracy, although we leave it to the reader to improve upon our current method.

We create a 24 separate regressors, which each predict Dst values some $n$ number of hours in advance, ranging from $n = 1$ to $n = 24$. For each time interval, we estimate regressors varying the number of previous input timesteps (between 1 and 24 hours of previous data) and data to include in the fit (Dst, electric field, and/or pressure). All regressors are trained on seven months of data

[2] In some instances this regressor may result in a lower MSE than previous examples in this section, due to the stochastic variability in MSE due to random forest hyperparameter fitting.

in 2010.[3] We then apply the regressor on Dst data in August 2010[4] and compute the MSE. The regressor with the lowest MSE becomes the main regressor for inferring Dst behavior $n$ hours in advance.

The combined array of 24 regressors can then be used to predict Dst all across the OMNI dataset. In Figure 5-18, we predict the storm behavior up to 24 hours in advance from the red point, spanning a range of data encapsulated within many of the examples within this section. We report forecasts for every hour up to 12 hours and, additionally, record the predicted Dst 24 hours in advance. Blue lines and scatterpoints indicate the predicted values. The blue shaded region indicates the 1 sigma uncertainty on the prediction from each random forest regressor. In this example, we naively estimate aleatoric uncertainty by recording the predictions from every decision tree in the random forest and computing the standard deviation of all predictions.[5]

In Figure 5-18, the forecast predicts the general trend of decreasing Dst over time, but fails to predict true Dst data within one standard deviation. We repeat this exercise on hundreds of different stretches of OMNI data containing a geomagnetic storm. Forecast accuracy varies significantly, with the model often capturing general Dst behavior but failing to capture minute changes in Dst over time. We present some particularly accurate examples in Figures 5-19 through 5-22. We note that this is a biased selection, and also present a poor example of Dst prediction in Figure 5-23.

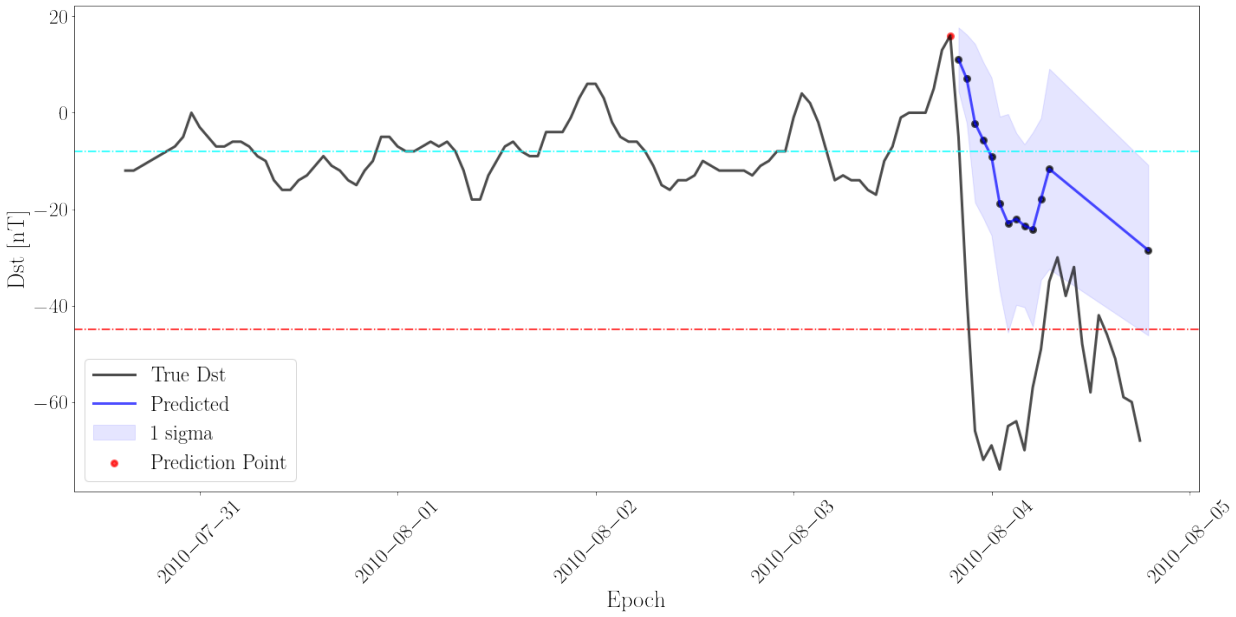



Figure 5-18: Forecasting Dst 24 hours out in August 2010.

We attempt to quantify our ability to forecast Dst through the creation of a P-P plot. To do this, we randomly select 500 times of OMNI data corresponding to the onset of a geomagnetic storm. We then predict Dst up to 24 hours in advance of each randomly selected start time, using our forecasting method. For each timestep, we then compare the predicted Dst to the true value. We record the percentile at which the true Dst value is recovered at in the prediction by computing

[3]Training on only OMNI data from 2010 is a potential source of bias. This tool can be improved by training on all OMNI data

[4]Again, another source of bias.

[5]See Section 6.1 for an additional note on uncertainty estimates with random forests.

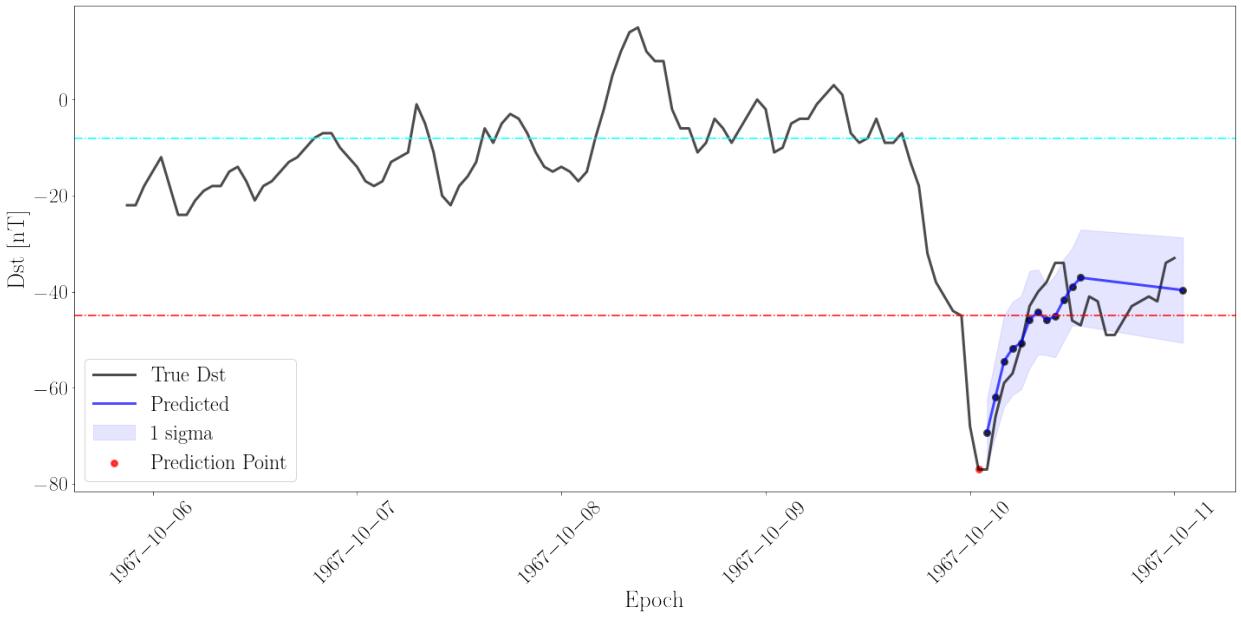



Figure 5-19: Forecasting Dst 24 hours out in October 1967.

the left-integrated percentile of the true value against a Gaussian with a mean of the predicted value and standard deviation corresponding to the random forest 1 sigma uncertainty. For all 500 examples, we record a percentile of recovery for each predicted timestep (every hour up to 12 hours, in addition to 24 hours). P-P plots capture a cumulative distribution function of these percentiles at each predicted time. The top panel of Figure 5-24 presents the P-P plot for 500 predictions at the onset of geomagnetic storms. A robust estimator would result in a perfect diagonal line, near the black dash–dotted line. While our estimators fail to produce a completely robust prediction, our predictions are not egregiously far off. In general, our forecasts result in underestimates of Dst at early times, and overestimates at late times. We repeat this analysis for 500 randomly selected times in the OMNI data set, regardless of the presence of a geomagnetic storm. The P-P plot for this analysis is presented in the bottom panel of Figure 5-24. This results in a similar trend to the top panel, although Dst is more frequently overestimated when a storm is not present. Both trends indicate that our forecast favors Dst values near the overall mean Dst of the OMNI data set, and struggles to predict outlier Dst behavior.

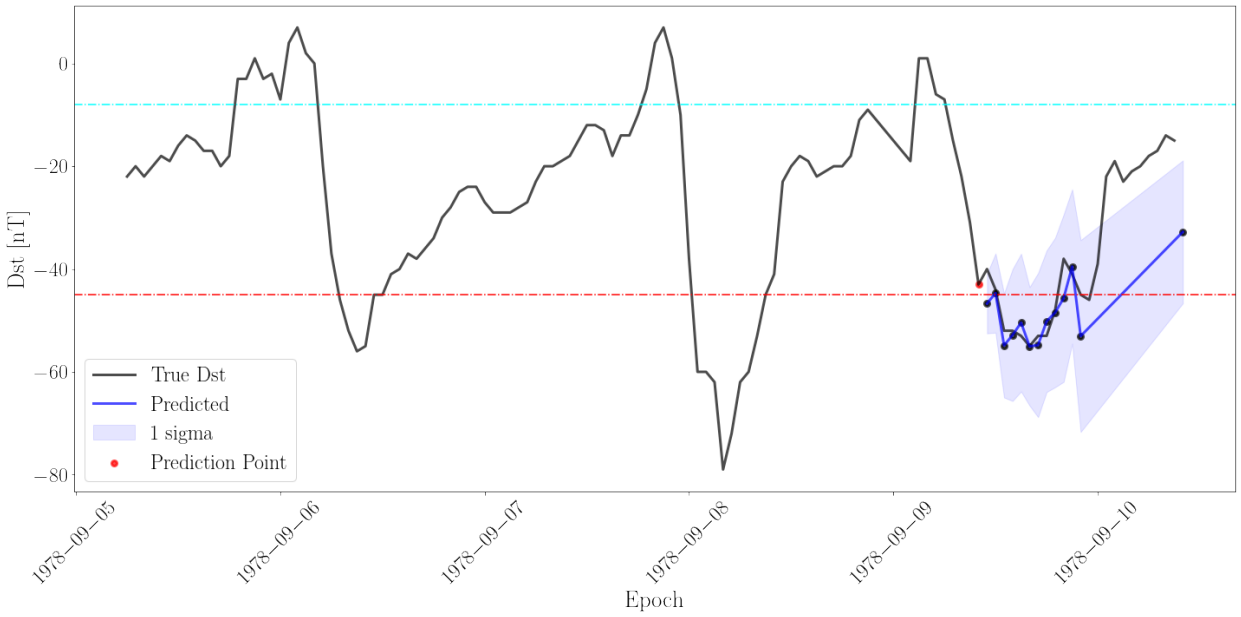



Figure 5-20: Forecasting Dst 24 hours out in September 1978.

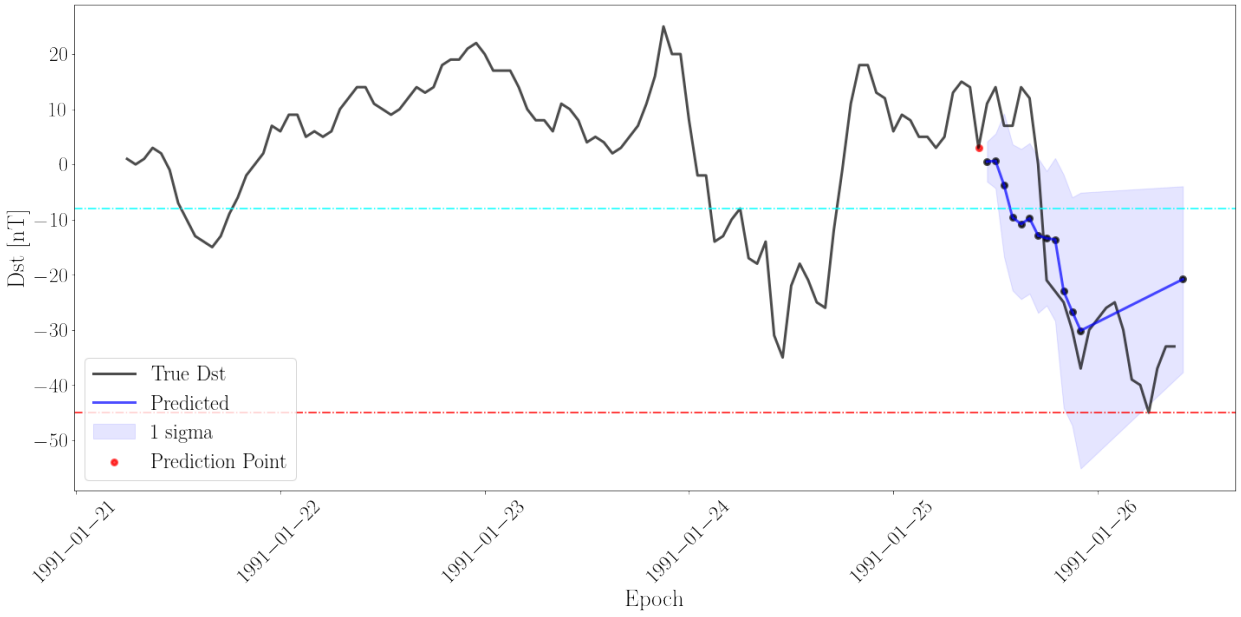



Figure 5-21: Forecasting Dst 24 hours out in January 1991.

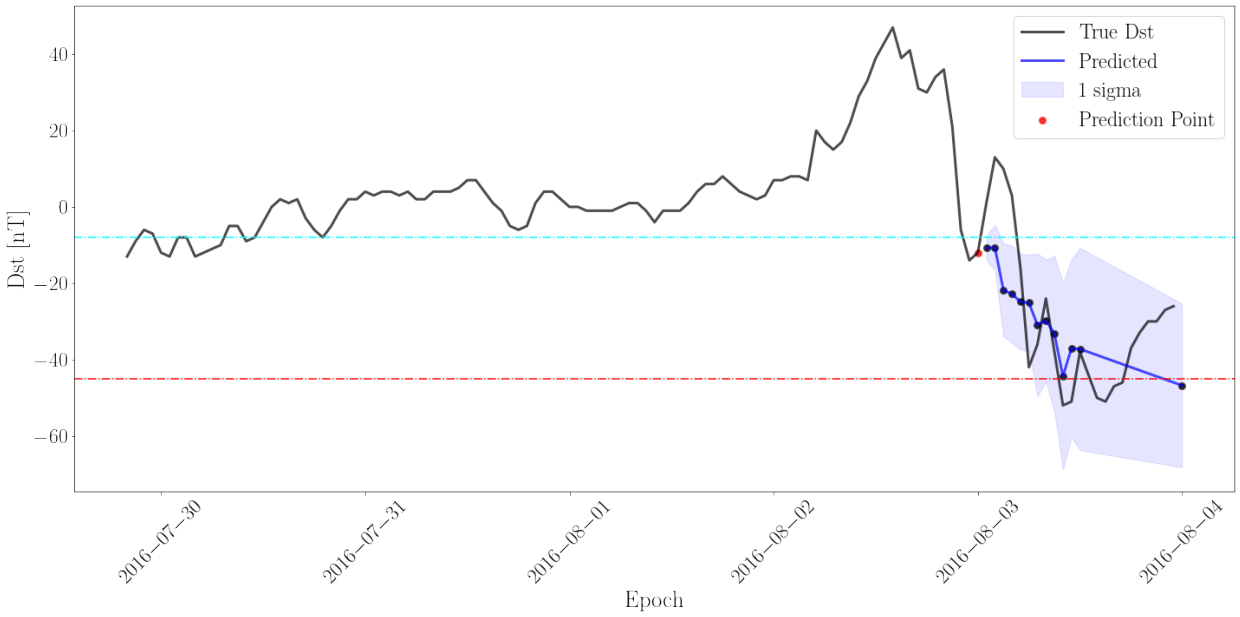



Figure 5-22: Forecasting Dst 24 hours out in August 2016.

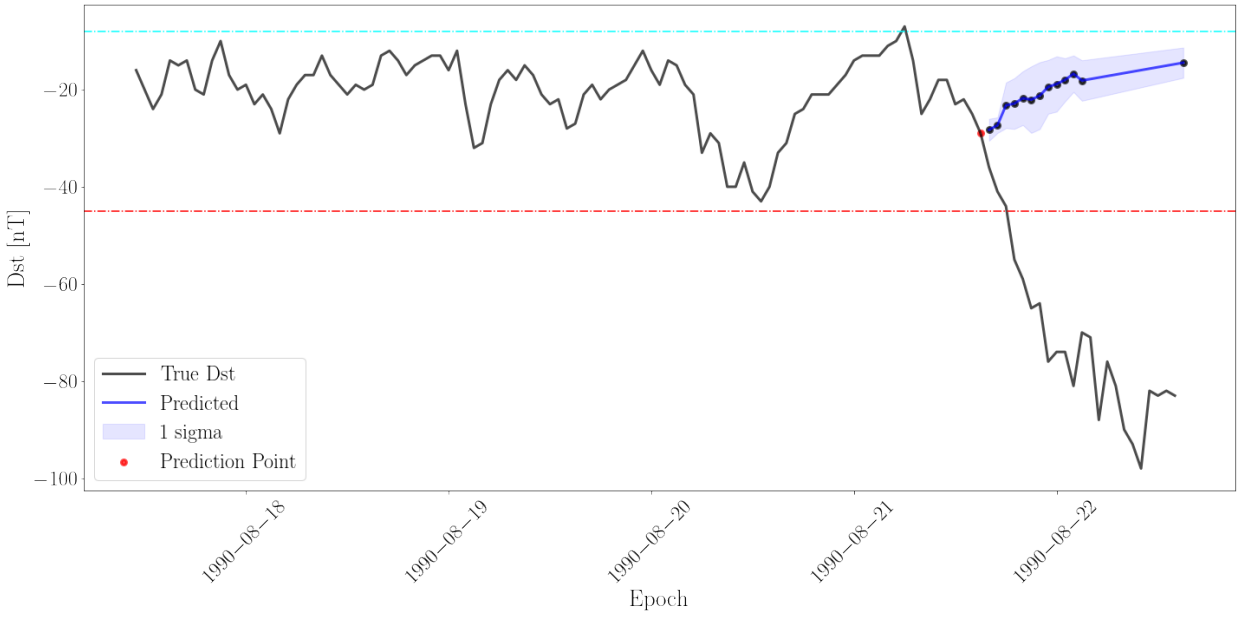
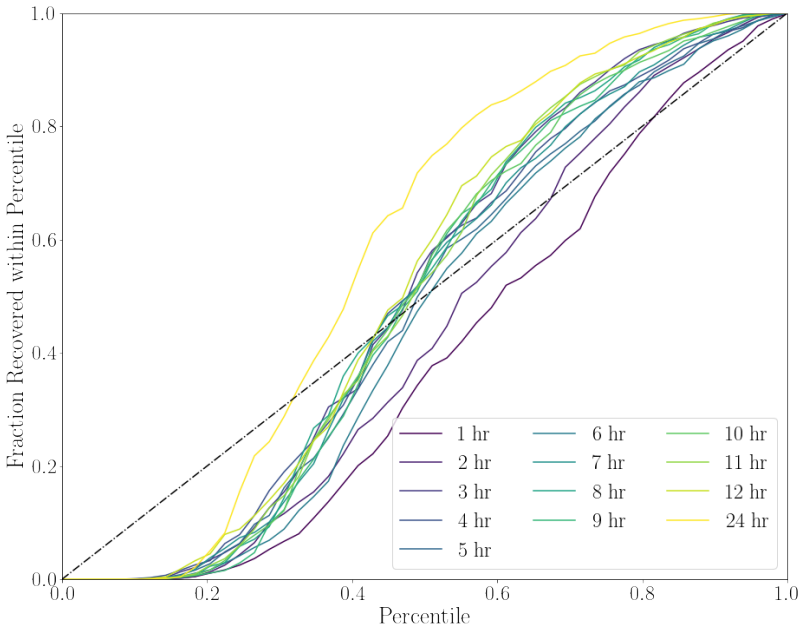



Figure 5-23: A poor attempt at forecasting Dst 24 hours out in August 1990.



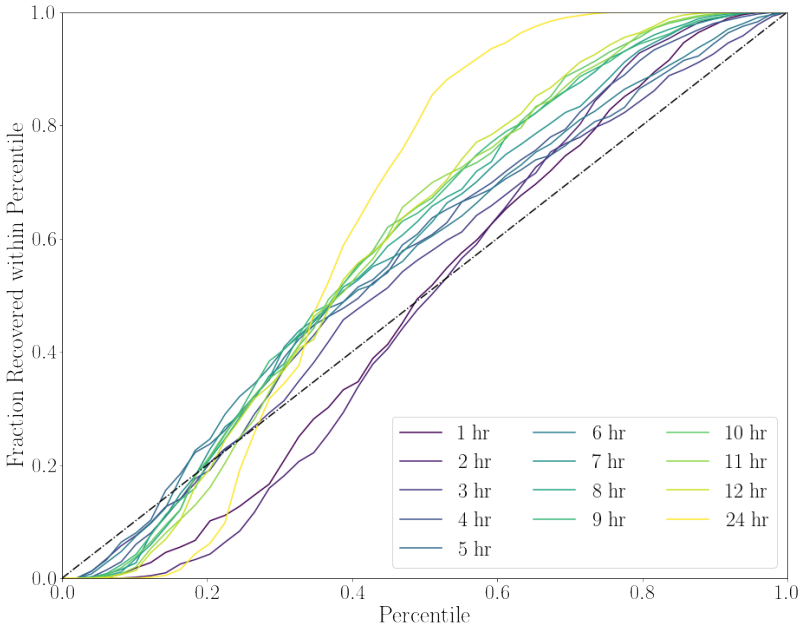




Figure 5-24: P-P plots of 500 Dst forecasts beginning at storm start times (top) or any time in the OMNI data (bottom).

## 5.5 Sliding Windows Conclusions

Sliding windows are a useful method for basic time series forecasting and show promise for predicting geomagnetic storm behavior from Dst alone. While we were unable to forecast Dst timeseries in an unbiased way, we anticipate that these methods could be built upon to accurately predict Dst.

# 6 Conclusions & Future Work

We have presented multiple attempts at inferring geomagnetic storm properties using a wide range of machine learning and statistical techniques. The methods presented in this work achieve varying success. We present numerous classification techniques in Section 3.1, none of which produce meaningful predictions of geomagnetic storm length. We note that Section 3.1 does not provide an exhaustive list of possible features to exploit for machine learning analysis.[6] In Section 4, we present a test case of using Gaussian Process Regression to extrapolate geomagnetic storm behavior; none of the Gaussian Process examples we explores were able to predict geomagnetic storms. We received the best results with Sliding Windows in Section 5, where Dst data itself was used to predict future Dst data. Unlike the classification data, we did not seek to infer any property of the storm, such as duration, but merely attempted to predict future Dst values. With more tuning, we anticipate that time slides may be a useful tool in predicting geomagnetic storms.

There are additional methods which may be useful in forecasting geomagnetic storms. For example, long short-term memory recurrent neural networks have shown some ability to predict storm behavior up to six hours in advance [6, 14].

## 6.1 A Note on Uncertainty

We do not provide uncertainty estimates on any value predicted by a random forest regressor throughout this text, except through a naive representation of aleatoric uncertainty in Section 5.4. There is no robust method to compute accuracy from a random forest regressor, although two approximate methods are commonly used in the literature. First, you can produce a crude estimate of uncertainty by recorded the output value by every decision tree in the random forest, as done in Section 5.4. Secondly, you can randomly train multiple random forest regressors, possibly on different combinations of training data, and produce a distribution of output values. Both fail to address underlying epistemic uncertainty in Dst measurements.

[6] We invite the reader to explore different possible input features, such as pressure or electric field.

## References


[1] URL: https://omniweb.gsfc.nasa.gov/.

[2] URL: http://wdc.kugi.kyoto-u.ac.jp/dstdir/.

[3] URL: https://isr-confluence.lanl.gov/display/NGPSSCI/Dst+duration+and+magnitude+prediction.

[4] Leo Breiman. "Random Forests." In: *Machine Learning* 45 (Jan. 2001), pp. 5–32. DOI: 10.1023/A:1010933404324.

[5] R. K. Burton, R. L. McPherron, and C. T. Russell. "An empirical relationship between interplanetary conditions and Dst". In: 80.31 (Nov. 1975), p. 4204. DOI: 10.1029/JA080i031p04204.

[6] M. A. Gruet et al. "Multiple-Hour-Ahead Forecast of the Dst Index Using a Combination of Long Short-Term Memory Neural Network and Gaussian Process". In: *Space Weather* 16.11 (Nov. 2018), pp. 1882–1896. DOI: 10.1029/2018SW001898.

[7] B. A. Larsen et al. *Extreme Values in the Radiation Belts: How Big and How Bad?* Report LA-UR-19-30361. Los Alamos National Laboratory, 2019.

[8] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[9] Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning.* Adaptive computation and machine learning. MIT Press, 2006, pp. I–XVIII, 1–248. ISBN: 026218253X.

[10] M. Temerin and Xinlin Li. "A new model for the prediction of Dst on the basis of the solar wind". In: *Journal of Geophysical Research (Space Physics)* 107.A12, 1472 (Dec. 2002), p. 1472. DOI: 10.1029/2001JA007532.

[11] M. Temerin and Xinlin Li. "Dst model for 1995-2002". In: *Journal of Geophysical Research (Space Physics)* 111.A4, A04221 (Apr. 2006), A04221. DOI: 10.1029/2005JA011257.

[12] H. L. Wei, S. A. Billings, and M. Balikhin. "Prediction of the Dst index using multiresolution wavelet models". In: *Journal of Geophysical Research (Space Physics)* 109.A7, A07212 (July 2004), A07212. DOI: 10.1029/2003JA010332.

[13] H. L. Wei et al. "Forecasting the geomagnetic activity of the Dst index using multiscale radial basis function networks". In: *Advances in Space Research* 40.12 (Jan. 2007), pp. 1863–1870. DOI: 10.1016/j.asr.2007.02.080.

[14] Wihayati, Hindriyanto Dwi Purnomo, and Suryasatriya Trihandaru. "Disturbance Storm Time Index Prediction using Long Short-Term Memory Machine Learning". In: *2021 4th International Conference of Computer and Informatics Engineering (IC2IE)*. 2021, pp. 311–316. DOI: 10.1109/IC2IE53219.2021.9649119.